Tellus **000**, 000–000 (0000) Printed 10 June 2010 (Tellus LaTeX style file v2.2)

# Quantifying the Benefit of A-SCOPE Data for Reducing Uncertainties in Terrestrial Carbon Fluxes in CCDAS

By T. Kaminski[1]*, M. Scholze [2] and S. Houweling[3],
[1] *FastOpt, Schanzenstr. 36, 20357 Hamburg, Germany;* [2] *QUEST – Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK;*[3] *Netherlands Institute for Space Research (SRON), Sorbonnelaan 2, 3584, CA Utrecht, The Netherlands*

(Manuscript received 20 January 2010, in final form 10 June 2010)

ABSTRACT

ESA's Earth Explorer candidate mission A-SCOPE aims at observing $CO_2$ from space with an active LIDAR instrument. This study employs quantitative network design techniques to investigate the benefit of A-SCOPE observations in a Carbon Cycle Data Assimilation System. The system links the observations to the terrestrial vegetation model BETHY via the fine resolution version of the atmospheric transport model TM3. In the modeling process chain the observations are used to reduce uncertainties in the values of BETHY's process parameters, and then the uncertainty in the process parameters is mapped forward to uncertainties in both in long-term net carbon flux and net primary productivity over three regions. A-SCOPE yields considerably better reductions in posterior uncertainties than the ground-based GLOBALVIEW station network. This is true for assimilating monthly mean values and instantaneous values, and it is true for two potential vertical weighting functions. The strength of the constraint through A-SCOPE observations is high over the range of observational uncertainties.

## 1 Introduction

$CO_2$ is the most important anthropogenic greenhouse gas, and the continued increase of atmospheric $CO_2$ is accepted to be the major reason for present, observed global warming. The increase of $CO_2$ is clearly of anthropogenic origin, but it is tempered by uptake from natural reservoirs. Therefore, understanding and predicting the cycling of this gas through natural and human-controlled systems is a matter of special importance. Despite considerable advances, major questions remain about the magnitude and distribution of present sources and sinks of this gas as well as their evolution, and their controlling mechanisms, especially their response to climate change. A clear requirement is the development of monitoring tools to ascertain the current sources and sinks and their changes. Such changes may occur as a result of climate change or of deliberate mitigation strategies. Hence, the monitoring of $CO_2$ and other greenhouse gases is immensely important, both for fundamental Earth System Science (as a necessary complement to global modeling), and for international climate policy.

A first awareness of global change began in the late 1950s when Charles David Keeling from the Scripps Institution of Oceanography, San Diego, developed a technique precise enough to detect the rise in atmospheric $CO_2$. Keeling's time series of atmospheric $CO_2$ concentration measurements from Mauna Loa (Keeling and Whorf, 2002) is an icon of contemporary environmental science. Major observational programs were subsequently put in place by various institutions from many different countries to create a global network, providing spatial gradients of $CO_2$ concentrations to constrain location and strength of $CO_2$ sources and sinks.

The current ground-based in-situ measurement network is mostly based on flask samples which are collected weekly to biweekly. The sampling stations are mainly located at remote sites to sample the $CO_2$ concentration of the marine boundary layer. A single flask measurement can be done with a high accuracy ($\approx$ 0.2 ppmv) such that a homogenized and gap-filled data product of this network like GLOBALVIEW (GLOBALVIEW-$CO_2$, 2004) reports uncertainties of 0.5 to 1 ppmv depending on the station location. However, the temporal (weekly to biweekly) and spatial (large gaps for instance in the tropics) resolution is fairly poor. Since the last few years new continuous atmospheric $CO_2$ observations have become available at some of these remote measurement stations to overcome the limitation in the temporal domain. Nevertheless, the network is still too sparse to quantify $CO_2$ sources and sinks on a regional to continental scale. This reflects the underdetermined nature of the inverse problem of inferring two-dimensional surface flux fields from point measurements (Kaminski and Heimann, 2001).

In addition to the ground-based flask and continuous station measurement system, there are a number of tall tower sites around the globe, which also provide in-situ continuous measurements of $CO_2$. These towers supply measurements of $CO_2$ in the continental boundary layer rather

* Corresponding author.
e-mail: Thomas.Kaminski@FastOpt.com

representative for regional to continental scale fluxes. For instance, NOAA is running around ten towers in the US, and the EU's CHIOTTO and then CarboEurope projects have set up around ten towers in Europe and Siberia. It is expected that this network will further expand in the future. Furthermore, there is an ever expanding network (see http://www.fluxnet.ornl.gov) providing local-scale direct flux observations via eddy covariance techniques.

Remote sensing of atmospheric $CO_2$ from space has the potential to deliver the data needed to substantially reduce the currently large uncertainties in the spatial and temporal distribution of $CO_2$ sources and sinks. Several sensitivity studies have evaluated the improvement in atmospheric inversion simulations of $CO_2$ that would be enabled by precise, global space-based integrated column $CO_2$ data. A pioneering study has been performed by Rayner and O'Brien (2001) who established the required precision for column-integrated $CO_2$ concentration data to be useful in constraining surface sources. Using an atmospheric synthesis inversion the required precision of monthly averaged (uniform weighting function) column data should be better than 2.5 ppmv (1.5 ppmv for oceanic coverage only) on a $8^\circ \times 10^\circ$ footprint for comparable performance with the existing surface network. They also reported that space-based column $CO_2$ observations with 1 ppmv precision were predicted to substantially reduce inferred $CO_2$ flux uncertainties of annual mean fluxes from greater than 1.2 GtC region$^{-1}$ year$^{-1}$ to less than 0.5 GtC region$^{-1}$ year$^{-1}$ when averaged over the annual cycle and for continent/ocean basin scale regions. Since then further studies (Rayner et al., 2002; Pak and Prather, 2001; Patra et al., 2003) have essentially confirmed the message that the overall precision on the measurements, including instrument uncertainties, noise, and uncertain atmospheric properties, needs to be better than 1% (or 3.6 ppmv) to provide a constraint on $CO_2$ fluxes comparable with the ground-based network.

More recently some studies have taken into account the characteristics and therefore potential benefits of different types of satellite instruments in synthetic inversion approaches. For example, Houweling et al. (2004) have distinguished between thermal infrared (AIRS) and near infrared (SCIAMACHY, OCO) spectrometers. The thermal infrared instrument AIRS has the advantage that it can measure the entire globe independent of day light or surface albedo, and thus has a relatively high number of measurements. Because of AIRS' independency on high surface albedo it is notably better performing over the oceans than SCIAMACHY. In contrast, OCO is able to measure in sun-glint mode over the oceans. A crucial factor is the ability to measure at low altitudes; therefore the near infrared instruments SCIAMACHY and OCO will certainly be more favorable as the thermal infrared instrument has a rather limited sensitivity to $CO_2$ near the surface. Their overall conclusion is that OCO will be the most promising satellite concept of those tested.

Miller et al. (2007) and Chevallier et al. (2007) specifically looked at the contribution of OCO column integrated $CO_2$ retrievals ($X_{CO2}$) to the reduction of uncertainties in the estimation of $CO_2$ sources and sinks. Both could show that, given the estimated error characteristics of the OCO instrument, OCO observations would significantly reduce the uncertainties of $CO_2$ surface fluxes, in the case of Chevallier et al. (2007) even at weekly timescale and grid point resolution of the underlying transport model ($2.5^\circ \times 3.75^\circ$) over land (reduction of 15-40% of prior uncertainties) and on monthly and basin-wide resolution over oceans (reduction of 20-40% of prior uncertainties). Two further recent studies also addressed the assessment of the OCO mission (Feng et al., 2009; Baker et al., 2008).

Unfortunately the launch of NASA's OCO mission in February 2009 failed. However, another satellite mission specifically aimed at measuring $CO_2$ from space, the GOSAT mission of the Japanese Aerospace Exploration Agency, was successfully launched in January 2009. GOSAT carries both a thermal and a near infrared spectrometer. The thermal spectrum provides similar information to what AIRS has been measuring, whereas the near infrared provides information about the total $CO_2$ column, which is more important for flux estimation. Chevallier et al. (2009) have quantified the potential of GOSAT data in a sensitivity study similar to the above mentioned studies. They found that GOSAT should significantly reduce uncertainties in $CO_2$ flux estimations over terrestrial vegetated areas at the scale of weeks and a few hundred kilometres, over the oceans improvements are only seen over larger scales (e.g. ocean basins and over a year).

The above mentioned approaches to convert atmospheric $CO_2$ concentrations into estimates of surface $CO_2$ fluxes are all based on 'top-down' inverse modeling of atmospheric transport. While this approach yields insights into the recent past and present, it cannot have predictive ability for the future. Future predictions, in contrast, are based on results from 'bottom-up' process-based model simulations. These simulations, however, lack the rigorous inclusion of the observational constraint.

An alternative method that is fully consistent with both the philosophy of inverse modeling, and the approach of predictive modeling employs techniques from variational data assimilation. In a first step both process parameters and initial conditions are estimated with the best possible accuracy using the best available observational constraints at the appropriate scale of the problem. A second (prediction or 'prognostic') step is then using not only standard modeling techniques employing the optimized parameters and initial conditions to arrive at a forecast, but also techniques of uncertainty propagation to estimate uncertainty ranges for the prediction (Scholze et al., 2007). This is a significant advance over current modeling techniques.

The Carbon Cycle Data Assimilation System (CCDAS) (Scholze, 2003; Rayner et al., 2005) is so far a unique example of the above outlined approach. It builds upon the study by Kaminski et al. (2002) who have used the seasonal cycle of atmospheric $CO_2$ to constrain a simplified terrestrial biosphere model. In CCDAS this simplified model is replaced by the more comprehensive, prognostic terrestrial biosphere model BETHY (Knorr, 2000; Knorr and Heimann, 2001). CCDAS uses a reduced version of BETHY which has no phenology scheme and no water balance. Instead it uses preoptimized leaf area index (LAI) and plant available soil moisture. Global vegetation is mapped onto 13 plant functional types (PFT). 57 control parameters affect the photosynthesis scheme, and both the autotrophic and heterotrophic respiration schemes. The assimilation of time series of atmospheric $CO_2$ flask data in CCDAS is controlled by a gradient algorithm, which searches the parameter space by iterative

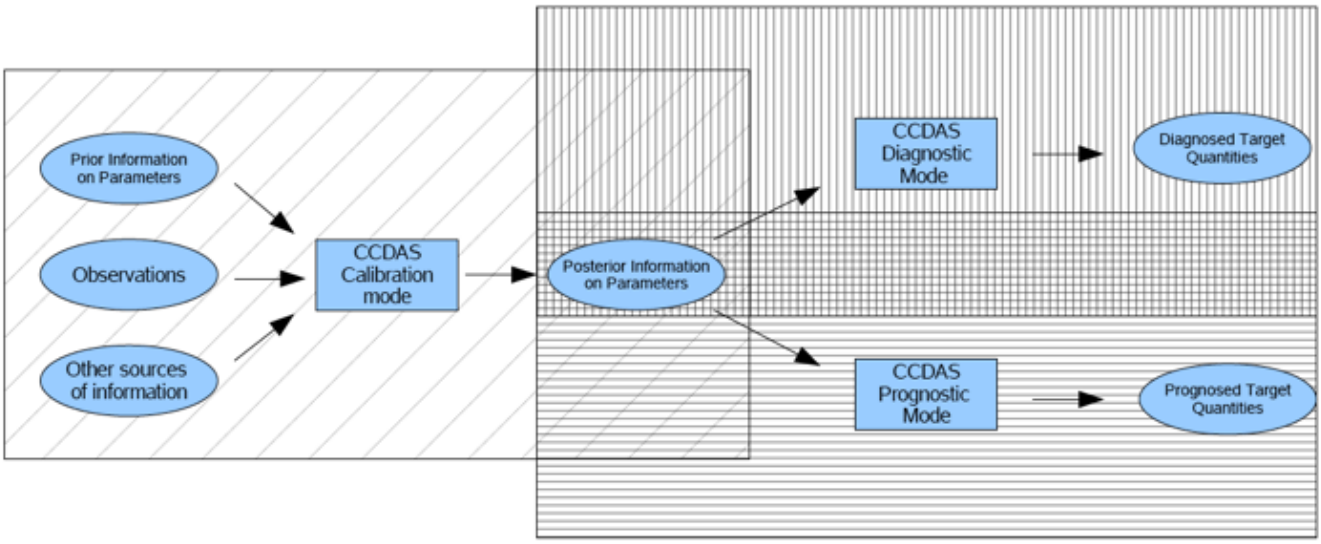


**Figure 1.** CCDAS two step procedure: Inverse step followed by diagnostic or prognostic step.

evaluation of a cost function and its gradient with respect to the parameters. The gradient information is provided efficiently by the model's adjoint. At the cost function minimum, an uncertainty of the estimated parameter set that is consistent with assumed observational and model uncertainties is approximated by the inverse of the function's full Hessian matrix, evaluated for the optimal parameter set up to machine precision. The calibration process, hence, delivers a set of optimized parameters, together with their uncertainties (see Figure 1).

CCDAS makes considerable use of derivative code, i.e. the adjoint code providing the gradient of the cost function, the Hessian code used to approximate parameter uncertainties, and Jacobian code to propagate these uncertainties forward. All derivative code is generated directly from the model's source code (Kaminski et al., 2003) by the automatic differentiation tool Transformation of Algorithms in Fortran (TAF, Giering and Kaminski (1998)). Since CCDAS has only 57 parameters, the evaluation of the full Hessian and the full Jacobian are computationally feasible.

In this study, we employ CCDAS to explore the benefit of the concept for ESA's Earth Explorer candidate mission A-SCOPE, which differs from the above mentioned concepts for observing $CO_2$ from space (AIRS, OCO, and GOSAT) in that it relies on an active LIDAR instrument. NASA is pursuing a similar concept with the ASCENDS mission (Michalak et al., 2008). The advantages of an active mission are that it does not require the sun as a light source, and can therefore provide both day and night, all-seasons and all latitude measurements and thus will provide an increase in the number of observations by a factor of two to three compared to passive missions. But more importantly, such an active mission concept provides a direct measurement of the atmospheric path and thus can assure the observation of the entire atmospheric column. This is an advantage over the OCO and GOSAT concept, which are particularly sensitive to the presence of aerosols, leading to potentially large gaps in regions with relatively persistent high levels of aerosols, such as some tropical regions (e.g. India), Southeast-Asia, or the Sahara. Two potential wavelengths, namely 1.6 micron and 2.0 micron, have been identified because of their high signal-to-noise ratio and favorable, near-uniform vertical weighting functions (see ESA (2008) for details on the A-SCOPE mission concept).

As a measure for the performance of A-SCOPE data we use the posterior uncertainty on regionally aggregated surface fluxes. Methodologically we are addressing a network design problem in a quantitative way. Quantitative network design was introduced to biogeochemistry by Rayner et al. (1996), who used an inverse model of the atmospheric transport to design surface networks, and it was also Rayner and O'Brien (2001) who first applied the approach to mission design. Kaminski et al. (2002) demonstrated the application of quantitative network design techniques for assimilating a synthetic flux measurement together with global atmospheric $CO_2$ samples and vegetation greenness approximated by AVHRR observations.

The remainder of this paper is organized as follows: Section 2 introduces quantitative network design methods and the extensions of CCDAS that were necessary to conduct the study. Section 3 describes and discusses the experiments that have been performed. Section 4 summarizes the main findings and draws conclusions.

## 2 Methodology

Methodologically, assessing the potential of a particular data stream in terms of quantifying a target quantity belongs to the class of so-called network design problems. This section gives a brief introduction to the mathematical formalism for quantitative network design, and then describes the extensions of CCDAS that were required to conduct the study.

### 2.1 Brief introduction to quantitative network design

Quantitative network design uses data assimilation systems and is, thus, closely linked to data assimilation. Hence, our introduction (following Kaminski and Rayner (2008)) starts off with the formalism behind Figure 1. In the formulation of the inverse problem it is convenient to quantify 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), the observational information by a PDF in the space of observations, and so on. Tarantola (1987) describes this probabilistic framework in detail and provides examples. Enting (2002) introduces the same framework with an exhaustive overview on applications to biogeochemistry.

If the input to the inverse problem can be characterized by Gaussian PDFs, the model that links control variables to observations is linear, and the model error follows a Gaussian PDF as well, then the posterior information is also quantified by a Gaussian PDF (see Tarantola (1987)). The mean of that PDF is given by:

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

and the covariance of its uncertainty is given by:

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

where $M$ denotes the Jacobian matrix of the model, $x_0$ and $C(x_0)$ the mean and covariance of the prior information's PDF. $d$ and $C(d)$ 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

$$C(d) = C(d_{\mathrm{obs}}) + C(d_{\mathrm{mod}}) \quad (3)$$

and by subtracting the mean model and observational errors from $Mx_0$ and $d$, respectively. Note that, in practice, these mean errors are usually difficult to assess.

It is easy to verify that $x$ (from eq. (1)) minimizes the cost function (the exponent of the Gaussian posterior PDF)

$$J(\tilde{x}) = \frac{1}{2}[(M\tilde{x} - d)^T C(d)^{-1} (M\tilde{x} - d) + (\tilde{x} - x_0)^T C(x_0)^{-1} (\tilde{x} - x_0)] \quad (4)$$

and that the Hessian matrix $H(\tilde{x})$ 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

$$C(x)^{-1} = H(x) \ . \quad (5)$$

If the model is non linear or any of the PDFs of the inputs are non Gaussian, eq. (1) and eq. (2) do not hold anymore. But we can still approximate the posterior PDF by a Gaussian with mean $x$ given by the minimum of eq. (4) (with the matrix $M$ generalized to the non linear model $M(\tilde{x})$) and covariance given by eq. (5).

In practice, any variational data assimilation system, e.g. in operational numerical weather prediction or oceanography, is based on eq. (4). The optimization mode of CCDAS uses an iterative procedure to minimize the cost function of eq. (4), which yields $x$, and computes $C(x)$ via eq. (5). 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. (4), and, hence, also to eq. (5), can be quantified by a separate term in the sum. In this formalism, the contribution of a synthetic data set (e.g. synthetic A-SCOPE observations) is to be handled as follows:

- The mean value is generated with the model itself, i.e. the equation $d = M(x)$ is applied, where $x$ is the best possible parameter value, taken from a minimization of eq. (4) for the existing observational network.
- The covariance of uncertainties (eq. (3)) is specified such as to reflect the expected characteristics of the observational products generated by the instrument and our ability to simulate them.

By construction of the synthetic data, their cost function contribution (eq. (4)) at the optimum, $x$, is zero but positive in the neighborhood of $x$. This means the synthetic data increase the curvature of the cost function. In mathematical terms, the curvature is expressed by the Hessian in eq. (5), which takes full account of the specified uncertainty in the synthetic data and their sensitivity to the model parameters. The effect of the synthetic data is a reduction of posterior parameter uncertainty.

The second step in Figure 1 is the estimation of a diagnostic or prognostic target quantity $y$, in our case some spatio-temporal mean carbon flux. The target quantity's PDF can be approximated by a Gaussian with mean

$$y = N(x) \quad (6)$$

and the covariance

$$C(y) = D' C(x) D'^T + C(y_{mod}) \ , \quad (7)$$

where $N$ is the model (in Figure 1 denoted as diagnostic/prognostic model) that maps the control variables onto the target quantity, $D'$ is its linearization around the mean of the posterior PDF of the control variables, also denoted as the Jacobian matrix of $N$, and $C(y_{mod})$ is the uncertainty in the model result 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 $C(x)$, and we omit the $C(y_{mod})$ contribution. If $N$ 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. (6) and eq. (7).

### 2.2 Including A-SCOPE data in CCDAS

One of the objectives of this study is to assess the data uncertainty for A-SCOPE that is required to achieve a given posterior uncertainty in a scalar target quantity $\sigma_y$. Since we will use a diagonal $C(d)$, with only two different entries for data over ocean and land, respectively denoted by $\sigma^2_{d,O}$ and $\sigma^2_{d,L}$, this can be done in a particularly efficient way.

Denoting the diagonal entries of $C(d)$ by $\sigma^2_{d,i}$, and the corresponding components of (the vector valued function) $M$ and $d$ respectively by $M_i$ and $d_i$, we rewrite the first term of eq. (4) in least squares form:

$$J(\tilde{x}) = \frac{1}{2}[\sum_i \frac{(M_i(\tilde{x}) - d_i)^2}{\sigma^2_{d,i}} + (\tilde{x} - x_0)^T C(x_0)^{-1} (\tilde{x} - x_0)]. \quad (8)$$

Taking second derivatives (one with respect to $\tilde{x}_k$ and one

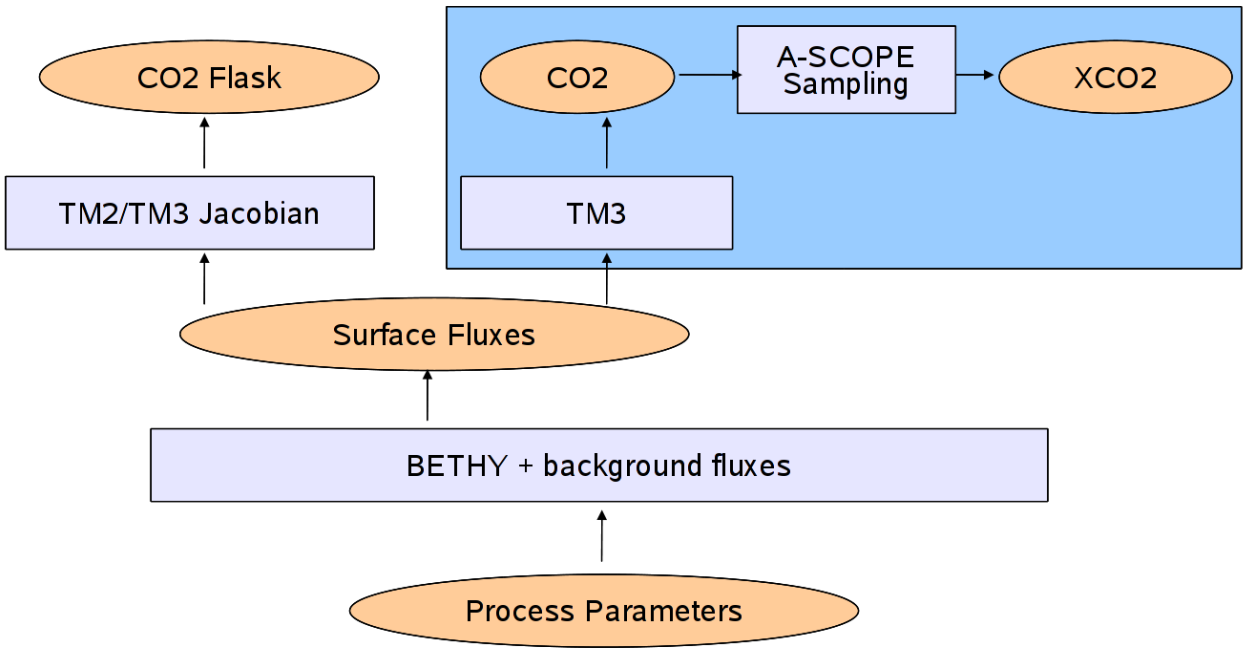


**Figure 2.** Modelling chain within CCDAS extended to include an A-SCOPE observation operator.

with respect to $\tilde{x}_l$) yields

$$H_{k,l}(\tilde{x}) = \frac{1}{2}\sum_i \frac{1}{\sigma_{d,i}^2}\frac{d^2(M_i(\tilde{x})-d_i)^2}{d\tilde{x}_k d\tilde{x}_l} + C_{k,l}(x_0)^{-1} \quad . \qquad (9)$$

Accumulating all summands for data over ocean and all summands for data over land, inserting into eq. (5) yields:

$$C(x)^{-1} = \frac{1}{\sigma_{d,O}^2}H_{A,O} + \frac{1}{\sigma_{d,L}^2}H_{A,L} + H_0 \qquad (10)$$

where the Hessian contribution of the prior is denoted by $H_0$ and the contribution by A-SCOPE is decomposed into two terms $H_{A,O}$ and $H_{A,L}$ determined by model characteristics and two factors depending on the data uncertainties over ocean and land.

Inserting in eq. (7) yields

$$\sigma_y^2 = D'(\frac{1}{\sigma_{d,O}^2}H_{A,O} + \frac{1}{\sigma_{d,L}^2}H_{A,L} + H_0)^{-1}D'^T + \sigma_{y,mod}^2 \qquad (11)$$

where, with our scalar target quantity, the Jacobian $D'$ takes the form of a row vector. Thanks to this decomposition we can precompute $H_{A,O}$ and $H_{A,L}$, such that a plot of $\sigma_y$ over $\sigma_d$ can be produced by pure matrix algebra without further CCDAS simulations.

For assessing the effect of A-SCOPE as an extension of the ground-based network, we can repeat the above algebra but starting from a cost function that has in addition to eq. (8) a third term representing the fit to the ground-based network. This situation is also covered by eq. (11), if we generalize the meaning of $H_0$ to represent the Hessians of all cost function contributions except the A-SCOPE term. In other words, in this situation $H_0$ denotes the Hessian when inverting against data from the ground-based network only.

In summary, the simulations require code for the computation of $H_0$, $H_{A,O}$, $H_{A,L}$ and $D'$. The computation of $H_{A,O}$ and $H_{A,L}$ requires the extension of CCDAS by the observation operator for A-SCOPE. Figure 2 shows the forward modelling chain of the extended CCDAS. The Biosphere Energy Transfer HYdrology scheme (BETHY, Knorr (2000); Knorr and Heimann (2001)) is used to simulate the surface fluxes of $CO_2$ from the terrestrial vegetation.

### *2.3 Representation of atmospheric transport*

Atmospheric $CO_2$ concentrations are modeled by the atmospheric transport model TM3 (Heimann and Körner, 2003), in its fine $4^\circ \times 5^\circ$ horizontal resolution, with 19 vertical $\sigma$ levels. As in Houweling et al. (2004), the model uses meteorological driving fields for the year 2000 as provided by Kalnay et al. (1996). Owing to the linearity of the atmospheric transport of $CO_2$, the vector $c$ of the changes in the total column $CO_2$ ($X_{CO2}$) at each observational location and time in response to a given flux field $f$ can be represented by its Jacobian matrix $A$:

$$c = Af \quad . \qquad (12)$$

The flux field $f$ is represented in the full $4^\circ \times 5^\circ$ resolution of the transport model and monthly temporal resolution. To compute one column of the Jacobian matrix corresponding to a given surface grid cell and month, the model is run with a unit emission in that grid cell and month (Enting, 2002). For each component of $c$, the simulated $X_{CO2}$ value corresponding to the observational location and time is recorded.

The above procedure would require one model run per grid cell and month in the period from the start of the CCDAS integration until the month in which the last $X_{CO2}$ ob-

servation takes place. To reduce the number of required model runs, we make two simplifications:

- For fluxes more than 4 years prior to a given $X_{CO2}$ observation we assume that their $CO_2$ emission is completely mixed within the global atmosphere. This means all columns in $A$ corresponding to fluxes more than 4 years ago contain a constant $a$ that quantifies the response in a completely mixed atmosphere (i.e. in a one box model).
- For fluxes less than 4 years prior to a given $X_{CO2}$ observation but not in the same month or the two months before, we assume that their $CO_2$ emission has the same effect as all other fluxes emitted in the same latitude band. This means all columns in $A$ corresponding to fluxes for a given month in said time span and in a given latitude band have the same response. Here, we use 8 latitude bands.

This means to simulate $c$ the response to the fluxes of three months $f_f$ is represented in full spatial resolution (full Jacobian) $A_f$, the response of the fluxes of 4 years (actually minus three months) $f_l$ is represented in latitudinal band resolution (latitudinal Jacobian) $A_l$, and the response to all previous years $f_g$ by a single number (global Jacobian) $A_g$, i.e.

$$c \approx A_f f_f + A_l f_l + A_g f_g \quad . \tag{13}$$

Figure 3 illustrates how the Jacobian is composed by blocks for full, latitudinal, and global Jacobians. The first row of blocks belongs to the earliest observations and the last row to the last observations. The first column belongs to the earliest fluxes and the last column to the last fluxes. Each block represents the impact of a particular month of fluxes on a particular month of concentrations. Concentrations are column integrated, taking the spectral weighting function and the orbit parameters into account.

We illustrate the computational savings through the matrix approximation in a sample calculation with 20 years of fluxes and observations in the last year, the set up used for this study. To provide the full Jacobian, $A_f$, we perform 12 sets of runs over three months each. The first covers the period from January to March, the second the period from February to April, etc. Each set consists of as many runs as we have grid cells, i.e. 72 × 46 (excluding the poles). We record the concentrations over the three months. The annual periodicity of the transport model's meteorological driving data is exploited in the following way: The simulation starting in December provides the response at the last concentration month (i.e. December) to fluxes in the same month, but it is also used to provide the response at the first concentration month (January) to fluxes one month prior. Without the periodicity we would need two additional sets of runs, for November and December in the year before the observations start. On the other hand, the sets of runs starting in November and December of the last year can be restricted to an integration period of two and one months, respectively. In total the computation of $A_f$ requires 39744 runs over three months each, i.e. 9936 years of transport model simulation. To provide the latitudinal Jacobian, $A_l$, we perform 12 sets of runs over 48 months each, where each set consists of as many runs as we use latitudinal bands in which all fluxes produce the same response, in our case 8. In total the computation of $A_l$ requires 96 runs over 48 months each, i.e. 384 years of transport model simulation. Finally, the global Jacobian, $A_g$, requires a single run over, say, five years. By contrast, without the approximation we would require 244 sets of runs, with integrations periods decreasing from 244 months to 1 month. Each set would consist of as many runs as we have grid cells, i.e. 72 × 46. This would produce a total of about 8.2 million years of transport model simulation.

The transport Jacobian is computed in two versions. The first version (instantaneous Jacobian) uses instantaneous samples at days 7, 14, 21, 28 of each month, and at 0.00 and 12.00 GMT. Note that for our assessment the modeled Jacobian does not need to match the exact date of the observation but only a meteorological situation typical for that time of the year. We can then assign every $X_{CO2}$ observation to the closest date in the record. The alternative version of the Jacobian uses monthly mean concentrations.

### *2.4 Uncertainties in observations and model*

For both Jacobian versions the specification of the data uncertainty is complicated by the fact that the simulated and observed quantities differ. The observed quantity is $X_{CO2}$ for a short interval in time and space (almost a point measurement) whereas the simulated quantity is a mean $X_{CO2}$ value. The monthly mean Jacobian simulates a mean over a horizontal grid cell and one month, whereas the instantaneous Jacobian simulates the mean over a horizontal grid cell and the model time step of 30 minutes.

Computing the difference between observed and simulated quantities brings in an additional source of uncertainty reflecting the error we make when transforming one quantity into the other. This error is called representation error (see, e.g, Heimann and Kaminski (1999)). We take this uncertainty into account by including an additional term $C(d_{\mathrm{rep}})$ in eq. (3) :

$$C(d) = C(d_{\mathrm{obs}}) + C(d_{\mathrm{rep}}) + C(d_{\mathrm{mod}}) \tag{14}$$

$C(d_{\mathrm{rep}})$ is hard to specify. We use a diagonal matrix, with the square of a constant $\sigma_{rep}$ on the diagonal. We derive a value based on the conservative assumption of $n$ point samples of a Gaussian distributed $X_{CO2}$ within the grid cell with standard deviation $\sigma_{het}$ ('het' standing for heterogeneity):

$$\sigma^2_{\mathrm{rep}} = \frac{\sigma^2_{\mathrm{het}}}{n} \tag{15}$$

$\sigma_{het}$ can, in principle, be observed. We use a conservative value of 3 ppmv for the total column, to reflect the fact that we also sample downstream of large fossil fuel emissions or over forests in the growing season. For the monthly mean Jacobian we use $n = 30$, which is about the (temporally and spatially varying) average sample size per horizontal grid cell and month, as derived from orbit simulations using MODIS cloud cover (Breon et al., 2009). For the instantaneous Jacobian $n = 1$, because we use each sample individually. A potential correlation between cloud cover and $X_{CO2}$ could be taken into account by a methodological refinement. First, $C(d_{\mathrm{rep}})$ would not be diagonal (uncertainty correlation across grid cells). Second, eq. (15) is too optimistic, because it is based on uncorrelated uncertainties of samples within the same grid cell. None of this is addressed in this study.

[t]

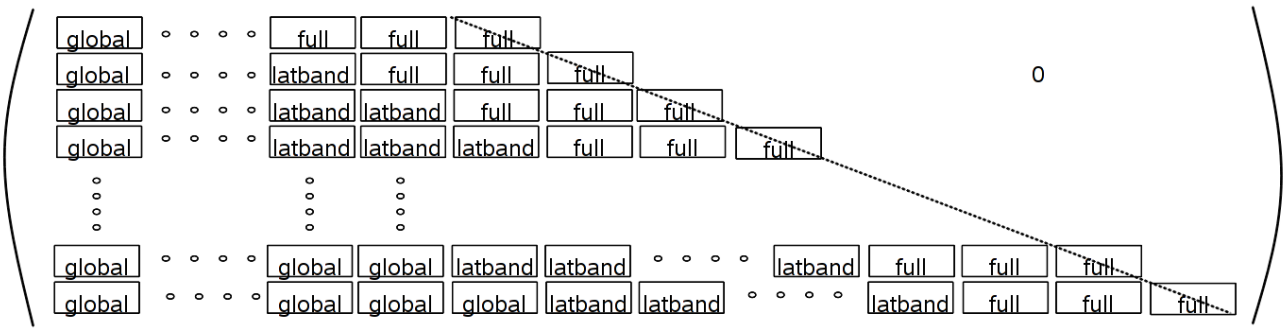


**Figure 3.** Structure of the Jacobian transport matrix.

For our base case with a vertical weighting function based on the 1.6 micron band (Ehret et al., 2008) we use observational uncertainties of 0.5 ppm over land and 1.5 ppm over the ocean respectively denoted $\sigma_{\mathrm{obs,L}}$ and $\sigma_{\mathrm{obs,O}}$. This corresponds to the target and threshold requirements of the A-SCOPE mission as given in the A-SCOPE Report for Assessment (ESA, 2008). For the 2.0 micron band we increase the observational uncertainties by a factor of two. Here again we are slightly optimistic by neglecting correlations in the observational uncertainties, i.e. by assuming only random errors. This is not too severe, because the A-SCOPE Report for Assessment (ESA, 2008), assumes only 10% of the observational uncertainty to be systematic. Also, calibration against ground measurements may help to build a model of the systematic error, e.g. with a uniform mean value (bias) plus a random component. The mean value can then be subtracted from the observations prior to assimilation. Only the random component then translates into a correlated uncertainty.

The uncertainty due to model error is also hard to specify. For the monthly mean Jacobian we use a diagonal form, with the square of a constant $\sigma_{mod}$ of 0.5 ppmv on the diagonal. This is probably conservative given that value we specify here has to be characteristic for the performance of a state-of-the-art model with a resolution of $4^\circ \times 5^\circ$. For the instantaneous Jacobian we also use a diagonal form but with a $\sigma_{mod}$ of 1.5 ppmv. We use this larger value for two reasons. First, the model error is larger for instantaneous values than for monthly mean values, where a fraction of the error cancels out. Second, the approximation of diagonal uncertainties is better for the monthly mean values than for instantaneous values, where correlated uncertainty from model error may play a larger role. Correlated uncertainties among a number of observations have the following effect on the cost function: They reduce the weight in the direction of the sum (average) of the observations and increase the weight in the direction of their differences. In our diagonal formulation we mimic the effect on the average by inflating the uncertainty. It is important to note that at this point we address only the residual model error for perfect parameter values and without any representation error, because the former is addressed by our method and the latter is accounted for by a separate term in eq. (14). Also, the positive correlations are partly compensated by processes involving mass conservation (e.g. within a carbon pool), which tend to create negative correlations among uncertainties.

For plotting the target uncertainty $\sigma_y$ over the observational uncertainty $\sigma_{\mathrm{obs}}$ we introduce a scaling factor $k$ for the observational uncertainty and combine eq. (11) with eq. (14):

$$\sigma_y^2 = D'\Big(\frac{1}{k\sigma_{\mathrm{obs,O}}^2 + \sigma_{\mathrm{rep}}^2 + \sigma_{\mathrm{mod}}^2} H_{A,O} + \frac{1}{k\sigma_{\mathrm{obs,L}}^2 + \sigma_{\mathrm{rep}}^2 + \sigma_{\mathrm{mod}}^2} H_{A,L} + H_0\Big)^{-1} D'^T + \sigma_{y,mod}^2 \quad . \quad (16)$$

$\sigma_{y,mod}$ refers to the uncertainty due to errors in the terrestrial model, and the first term specifies the uncertainty in case we had a perfect model for $y$. $\sigma_{y,mod}$ is, of course, strongly dependent on the model and difficult to assess. In order not to mask the assessment of A-SCOPE through a rather arbitrary assumption on $\sigma_{y,mod}$, we don't include it in the default computation of $\sigma_y$. We do this consistently throughout all experiments, i. e. our A-SCOPE assessment and our benchmark, the ground-based station network. For both the A-SCOPE assessment and the benchmark exactly the same model error is to be used in eq. (16). Owing to the simple dependency of $\sigma_y$ on $\sigma_{y,mod}$, one can easily combine it with the uncertainties we provide. We show, however, an example calculation that estimates the model error from an ensemble of terrestrial biosphere models. Cramer et al. (2001) (Fig. 5) compare NPP and NEP simulated by six terrestrial biosphere models. Both quantities depend on the process representations in the individual models. For the 1990s the global NPP of the six models spans a range of about 15 GtC per year, while NEP spans a range of about 1.5 GtC, and with one outlier removed below 0.5 GtC. For the definition of relative ranges it makes sense to refer to global NPP also for the NEP range, because NEP is the difference between two large fluxes. The ranges relative to 60 GtC (a typical value for global NPP, our model's global

NPP is 64.9 GtC) are then 25 % for NPP and 2.5 % for NEP. For our three regions the 25 % of NPP are conservative: The global map in Fig 2. of Cramer et al. (1999) shows a lower relative range (in this study for the 16 models of the Potsdam NPP intercomparison) over these regions, but a larger relative range over Africa. This is also confirmed by Fig. 1 of Kicklighter et al. (1999) who show the NPP range spanned by 90% of the same models over latitude. Over our twenty year integration period the long-term NPP average per year is 2.6 GtC over Europe, 6.5 GtC over Russia, and 6.1 GtC over Brazil. Associating the above derived relative ranges with a +/- one standard deviation interval (again a conservative assumption), yields respectively for the three regions per year 0.3 GtC, 0.8 GtC, 0.8 GtC for NPP and 0.03 GtC, 0.08 GtC, 0.08 GtC for NEP. Our computational example is based on spreads in model simulations representing the state-of-the-art around the year 2000. We can hope that these spreads decrease with progress in terrestrial modelling, and with systematic model calibration against observations. Our example is also conservative in that the spreads include the parametric uncertainty (caused by wrong values of the process parameters), a source of uncertainty that we explicitly specify in eq. (16).

## 3 Experiments

This section describes the four experiments that are performed with the extended CCDAS. All experiments are run with the CCDAS configuration determined by the optimized parameter values (see Table 1) from Scholze et al. (2007), who use 41 sampling sites from the (GLOBALVIEW-$CO_2$, 2004) network. The total data uncertainty $\sigma_d$ is 1.08 ppmv per observation on average. For a detailed description of the parameters we refer to Rayner et al. (2005). The simulation period covers 20 years with the ground-based network sampling over the entire period and A-SCOPE only in the final year. The Hessian for the ground-based network ($H_0$ in eq. (10), eq. (11), and eq. (16)) is taken from Scholze et al. (2007). The same holds for the Jacobian mapping parameter uncertainty onto flux uncertainty ($D'$ in eq. (11) and eq. (16)). The experiments use the following seven target quantities: net carbon flux (NEP) and net primary productivity (NPP) over Europe, Russia, and Brazil, as well as global NEP for consistency checks.

The experimental setups are as follows:

1) **Base experiment**: This experiment applies vertical weighting for the 1.6 micron band (see Table 2) to the transport Jacobian with monthly mean concentrations and assumes global coverage with observational uncertainties of 0.5 ppmv over land and 1.5 ppmv over ocean plus an uncertainty of 0.5 ppmv reflecting model error. The first simulation (Case 1) uses both the ground-based flask sampling network and A-SCOPE, the second simulation (Case 2, base case) runs without the ground-based flask station network and only A-SCOPE sampling, and the third simulation (Case 3) with the ground-based station network only. This experiment allows us to assess the benefit of A-SCOPE and the ground-based flask sampling network separately and in conjunction. Cases 2* and 2** illustrate the effect of including the uncertainty due to model error ($\sigma_{y,mod}$ term in eq. (16)).

**Table 1.** CCDAS parameter values used for the experiments. Units are $V_{max}$: $\mu$mol($CO_2$)$m^{-2}s^{-1}$, $a_{J,T}$: (deg C)$^{-1}$, $a_{\Gamma,T}$: $\mu$mol($CO_2$)mol(air)$^{-1}$(deg C)$^{-1}$, activation energies E: J/mol, $\tau_f$: years, offset: ppm, all others unit-less. Uncertainties are in percentage except for log-normally distributed parameters for which a range is given. Uncertainties represent one standard deviation.

| Number | Parameter | value | uncertainty |
|---|---|---|---|
| 1 | $V_{\max}^{25}$(TrEv) | 57.6 | 20 |
| 2 | $V_{\max}^{25}$(TrDec) | 108.5 | 20 |
| 3 | $V_{\max}^{25}$(TmpEv) | 40.7 | 20 |
| 4 | $V_{\max}^{25}$(TmpDec) | 51.2 | 20 |
| 5 | $V_{\max}^{25}$(EvCn) | 26.0 | 20 |
| 6 | $V_{\max}^{25}$(DecCn) | 119.1 | 20 |
| 7 | $V_{\max}^{25}$(EvShr) | 130.4 | 20 |
| 8 | $V_{\max}^{25}$(DecShr) | 137.0 | 20 |
| 9 | $V_{\max}^{25}$(C3Gr) | 11.4 | 20 |
| 10 | $V_{\max}^{25}$(C4Gr) | 0.4 | 20 |
| 11 | $V_{\max}^{25}$(Tund) | 35.6 | 20 |
| 12 | $V_{\max}^{25}$(Wetl) | 19.2 | 20 |
| 13 | $V_{\max}^{25}$(Crop) | 95.6 | 20 |
| 14 | $a_{\mathrm{J,V}}$(TrEv) | 1.92 | 5 |
| 15 | $a_{\mathrm{J,V}}$(TrDec) | 1.99 | 5 |
| 16 | $a_{\mathrm{J,V}}$(TmpEv) | 2.0 | 5 |
| 17 | $a_{\mathrm{J,V}}$(TmpDec) | 2.0 | 5 |
| 18 | $a_{\mathrm{J,V}}$(EvCn) | 1.79 | 5 |
| 19 | $a_{\mathrm{J,V}}$(DecCn) | 1.82 | 5 |
| 20 | $a_{\mathrm{J,V}}$(EvShr) | 1.97 | 5 |
| 21 | $a_{\mathrm{J,V}}$(DecShr) | 1.66 | 5 |
| 22 | $a_{\mathrm{J,V}}$(C3Gr) | 1.88 | 5 |
| 23 | $a_{\mathrm{J,V}}$(C4Gr) | 9.9 | 5 |
| 24 | $a_{\mathrm{J,V}}$(Tund) | 1.86 | 5 |
| 25 | $a_{\mathrm{J,V}}$(Wetl) | 1.84 | 5 |
| 26 | $a_{\mathrm{J,V}}$(Crop) | 1.92 | 5 |
| 27 | $\alpha_q$ | 0.34 | 5 |
| 28 | $\alpha_i$ | 0.04 | 5 |
| 29 | $K_{\mathrm{C}}^{25}$ | $445\times10^{-6}$ | 5 |
| 30 | $K_{\mathrm{O}}^{25}$ | 0.33 | 5 |
| 31 | $a_{\Gamma,T}$ | 1.45 | 5 |
| 32 | $E_{\mathrm{K_O}}$ | 36218 | 5 |
| 33 | $E_{\mathrm{K_C}}$ | 58637 | 5 |
| 34 | $E_{\mathrm{V_{max}}}$ | 62045 | 5 |
| 35 | $E_{\mathrm{k}}$ | 50592 | 5 |
| 36 | $E_{\mathrm{R_d}}$ | 42023 | 5 |
| 37 | $f_{\mathrm{R,leaf}}$ | 0.26 | 25 |
| 38 | $f_{\mathrm{R,growth}}$ | 1.12 | 5 |
| 39 | $f_S$ | 0.43 | -0.1; +0.2 |
| 40 | $\kappa$ | 0.59 | -0.9; +9.0 |
| 41 | $Q_{10,\mathrm{f}}$ | 2.00 | -0.5; +0.75 |
| 42 | $Q_{10,\mathrm{s}}$ | 1.31 | -0.5; +0.75 |
| 43 | $\tau_f$ | 6.8 | -1.0; +3.0 |
| 44 | $\beta$(TrEv) | 1.33 | 25 |
| 45 | $\beta$(TrDec) | 1.01 | 25 |
| 46 | $\beta$(TmpEv) | 1.23 | 25 |
| 47 | $\beta$(TmpDec) | 2.55 | 25 |
| 48 | $\beta$(EvCn) | 0.74 | 25 |
| 49 | $\beta$(DecCn) | 1.77 | 25 |
| 50 | $\beta$(EvShr) | 1.21 | 25 |
| 51 | $\beta$(DecShr) | 0.20 | 25 |
| 52 | $\beta$(C3Gr) | 0.77 | 25 |
| 53 | $\beta$(C4Gr) | 0.76 | 25 |
| 54 | $\beta$(Tund) | 1.14 | 25 |
| 55 | $\beta$(Wetl) | 0.56 | 25 |
| 56 | $\beta$(Crop) | 3.26 | 25 |
| 57 | offset | 336.3 | 0.3 |

**Table 2.** Input quantities for vertical weighting for calculation of $X_{CO2}$ concentration on TM3 fine grid configuration with 19 $\sigma$ levels.

| level | press (Pa) | weight 1.6 micron | weight 2.0 micron |
|---|---|---|---|
| 1 | 97908.00 | 0.050 | 0.107 |
| 2 | 96637.70 | 0.051 | 0.106 |
| 3 | 94895.50 | 0.051 | 0.105 |
| 4 | 92748.40 | 0.052 | 0.103 |
| 5 | 88410.40 | 0.054 | 0.100 |
| 6 | 80892.20 | 0.057 | 0.094 |
| 7 | 70973.50 | 0.062 | 0.085 |
| 8 | 59045.40 | 0.068 | 0.073 |
| 9 | 46180.60 | 0.075 | 0.058 |
| 10 | 33827.50 | 0.079 | 0.043 |
| 11 | 25439.80 | 0.077 | 0.031 |
| 12 | 20697.90 | 0.071 | 0.025 |
| 13 | 16579.90 | 0.062 | 0.019 |
| 14 | 13072.40 | 0.052 | 0.015 |
| 15 | 10134.70 | 0.043 | 0.012 |
| 16 | 7708.66 | 0.035 | 0.009 |
| 17 | 5728.36 | 0.028 | 0.007 |
| 18 | 3556.67 | 0.021 | 0.005 |
| 19 | 1136.03 | 0.012 | 0.003 |

2) **Sensitivity to temporal and horizontal sampling**: This experiment repeats Experiment 1, Case 2, but with different sampling. Based on the above-mentioned A-SCOPE orbit specification by Breon et al. (2009) it assumes sampling of instantaneous concentrations, for which the sensitivity uses the nearest location, day and time of day in the instantaneous transport Jacobian.
3) **Sensitivity to vertical weighting**: This experiment repeats Experiment 1, Case 2, but with the 2.0 micron band vertical weighting (see Table 2) instead of the 1.6 micron band weighting and with observational uncertainties increased by a factor of two.
4) **Sensitivity to data uncertainties**: This experiment repeats Experiment 1, Case 2 with a joint scaling factor for the two observational uncertainties over land and ocean in eq. (10) and samples 25 values of this scaling factor.

Table 3 lists the prior and posterior uncertainties for Experiments 1 to 3. In Experiment 1, Case 2, which is our base case, the posterior uncertainties are strongly reduced compared to the prior uncertainties. This is the case for both NPP and NEP over all three regions. The strongest reduction occurs for global NEP, with a posterior uncertainty of 0.013 GtC per year. For a consistency check we can calculate a lower bound for that value by thinking of a well-mixed atmosphere that is sampled at the end of the integration period by all A-SCOPE observations. The number of observations on the 72 by 48 grid over 12 months is about 40000, and the average data uncertainty about 1 ppmv, which yields an uncertainty of 0.005 ppmv. We can infer, as a single unknown, a 20 year global mean NEP by inverting the box model. This is particularly easy if we assume that the initial concentration (parameter 57) is known and neglect the contribution of the prior uncertainty in eq. (2). Using a conversion factor of $20 \cdot 0.5$ year ppmv per GtC, we end up with a value of 0.0005 GtC per year. This is consistent with the experiment's posterior uncertainty of 0.013 for global NEP, which is well above this lower bound.

Note that posterior uncertainties for A-SCOPE derived

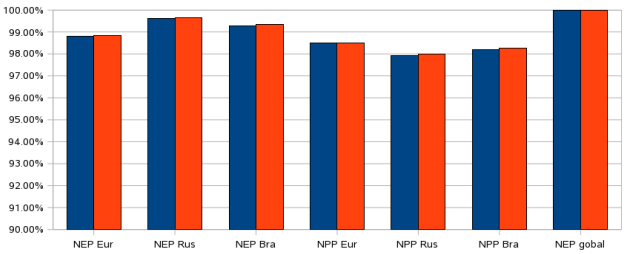


**Figure 4.** Uncertainty reduction (relative to prior uncertainty) for A-SCOPE only in blue (case 2). For comparison red bars show uncertainty reduction for case 1

by CCDAS are generally lower than those of assessments for the OCO mission by classical transport inversions (Miller et al., 2007; Chevallier et al., 2007; Baker et al., 2008) or Kalman filters (Feng et al., 2009). Their values, however, refer to considerably shorter temporal averaging periods than our twenty years. Extending the averaging period typically reduces the uncertainties owing to negative correlations in uncertainties along the temporal axis. This is well known in flux inversions but also holds in CCDAS (Scholze et al., 2007). A detailed attribution of the differences in posterior uncertainties to factors such as the averaging period, the additional constraint through our terrestrial model, and the mission concept is far beyond the scope of the present study.

The benefit from the observational constraint is limited by the uncertainty from errors in the terrestrial model. The two bottom rows of Table 3 (labeled Experiment 1, Cases 2* and 2**) illustrate the effect of including a $\sigma_{y,mod}$ term in eq. (16) for the calculation of Experiment 1, Case 2. For Case 2* we use the values for the example calculation for $\sigma_{y,mod}$ based on model spread representing the state-of-the-art around the year 2000 (see section 2.2). Note that this calculation produces the total uncertainty in model output, i.e. it also includes the parametric uncertainty, which our method already accounts for. Hence, assigning this total uncertainty to $\sigma_{y,mod}$ (the fraction of uncertainty in model output that is not produced by parametric uncertainty) is extremely cautious. This is also indicated by the large uncertainties produced by the prior parameter uncertainties (row 1). Meanwhile a number of benchmarking activities (e.g., Randerson et al. (2009); Cadule et al. (2010)) are aiming at separating realistic from unrealistic process formulations in terrestrial biosphere models. Such activities will drastically narrow down $\sigma_{y,mod}$: To reflect the anticipated progress, Case 2** uses a $\sigma_{y,mod}$ of 10% of the above-calculated total uncertainty.

The reduction of uncertainty relative to the prior uncertainty quantifies the strength of the observational constraint on the respective target quantity. It is shown for all target quantities in Figure 4. For each of the regions, NEP is better observed than NPP. The comparison between Cases 2 (black bars) and 1 (grey bars) shows that adding the station network yields only slight improvement. The constraint by the station network alone is about a factor 20-40 weaker than the constraint by A-SCOPE.

Uncertainty reductions for Experiments 2 and 3 are in the same range as for Case 2 of Experiment 1. This means the good performance of A-SCOPE is robust against the horizontal and temporal averaging in the Jacobian and a

**Table 3.** Prior and posterior uncertainties for all experiments in GtC/year.

| Experiment | Case | NEP Eur | NEP Rus | NEP Bra | NPP Eur | NPP Rus | NPP Bra | NEP gobal |
|---|---|---|---|---|---|---|---|---|
| prior | | 0.4500 | 1.5000 | 1.1000 | 0.6600 | 1.1000 | 4.9000 | 52.0000 |
| 1 | 1 | 0.0052 | 0.0054 | 0.0072 | 0.0100 | 0.0220 | 0.0850 | 0.0092 |
| 1 | 2 | 0.0054 | 0.0057 | 0.0081 | 0.0100 | 0.0230 | 0.0880 | 0.0130 |
| 1 | 3 | 0.1700 | 0.1100 | 0.2800 | 0.3300 | 0.4900 | 1.2000 | 0.1200 |
| 2 | | 0.0059 | 0.0050 | 0.0090 | 0.0100 | 0.0220 | 0.1100 | 0.0100 |
| 3 | | 0.0042 | 0.0042 | 0.0060 | 0.0082 | 0.0190 | 0.0750 | 0.0130 |
| 1 | 2 * | 0.0329 | 0.0814 | 0.0767 | 0.3252 | 0.8128 | 0.7676 | 0.8114 |
| 1 | 2 ** | 0.0063 | 0.0099 | 0.0111 | 0.0340 | 0.0844 | 0.1164 | 0.0822 |

change of the spectral band with the associated change in weighting function and observational uncertainty. In Experiment 2 the inflated uncertainty for $\sigma_{mod}$ in eq. (14) meant to compensate for challenges in modelling instantaneous samples as well as possible correlations in uncertainties due to model error did not degrade the performance of A-SCOPE.

Experiment 4 enables us to plot the posterior uncertainty in all seven target quantities over the scaling factor. Figure 5 shows a modest sensitivity of the target uncertainty to the scaling. This is because the scaling is deliberately restricted to the observational uncertainty while uncertainties reflecting model and representation error are kept constant.

## 4 Conclusions

The present study investigated the benefit of A-SCOPE observations in a CCDAS that links the terrestrial vegetation model BETHY (Knorr, 1997; 2000) to observations of $CO_2$ total column content via the fine resolution version of the atmospheric transport model TM3 (Heimann and Körner, 2003). In the modelling process chain the observations are used to reduce uncertainties in the values of BETHY's process parameters, and then the uncertainty in the process parameters is mapped forward to uncertainties in both net carbon fluxes (NEP) and net primary productivity (NPP) over three regions. Note that traditional transport inversions cannot handle target quantities other than NEP.

For the assessment, other sources of carbon dioxide (the so-called background fluxes) such as fossil fuel emissions, land use change fluxes and exchange fluxes with the ocean were prescribed to fixed values without uncertainty. We are thus likely to over estimate the A-SCOPE constraint on the terrestrial process parameters and, hence, also on the calculated fluxes.

A-SCOPE yields considerably better reductions in posterior uncertainties than the ground-based GLOBALVIEW station network used by Scholze et al. (2007). This is true for assimilating monthly mean values and instantaneous values, and it is true for both the 1.6 micron band and the 2.0 micron band vertical weighting function. The strength of the constraint through A-SCOPE observations is high over the range of observational uncertainties from 0.05 to 1.25 ppmv over land and from 0.15 to 3.75 ppmv over ocean. A potential A-SCOPE mission would, thus, have a major impact on our understanding of the global carbon cycle and narrow down the currently large uncertainties in future climate simulations owing to the climate-carbon cycle feedback (Friedlingstein et al., 2006).

The reasons for the strong constraint lie in the real global coverage and the much larger number of observations compared to the GLOBALVIEW station network. In contrast to pure transport inversions, the CCDAS approach exploits the powerful constraint provided by the terrestrial process formulations within the vegetation model. The model classifies global vegetation into 13 plant functional types (PFTs) grouped according to the plants' morphology, physiology, phenology as well as bioclimatic limits. Each PFT has its own set of process parameters and provides a strong link between the various regions of occurrence. Hence, an observation over one region can help to constrain fluxes over another region. For a model with more PFTs, or spatially varying parameter values, the observational constraint would be weaker.

Similar to pure transport inversions our study results also depend on the assumptions on uncertainties that reflect model and representation errors. This study neglected correlated uncertainties for the observations and for the sampling of the horizontal mean $X_{CO2}$ concentrations over a TM3 grid cell. The A-SCOPE Report for Assessment (ESA, 2008) specifies a systematic error contribution of as low as 10% to the total observational uncertainty. It is desirable to stay with a mission design that assures a low and uniform systematic error. This is because calibration against ground measurements may help to build a model of the systematic error, e.g. with a uniform mean value plus a random component. The mean value can then be subtracted from the observations prior to assimilation and only the random component enters the inversion in the form of a correlated uncertainty.

Our estimate of the uncertainty that reflects representation error is based on independent sampling of the $X_{CO2}$ concentration within a horizontal transport model grid cell of $4^\circ \times 5^\circ$ and an ad hoc assumed variability of 3 ppmv. This yields a relatively low representation error. It would be desirable to derive an improved estimate reflecting observed small-scale variability. Traditional transport inversions would also benefit in the same way from a better quantification of representation error.

Any comparison based on published work with classical transport inversions is difficult, because differences in posterior flux uncertainties are affected by a number of factors such as the constraint through the terrestrial model as mentioned above, but also through differences in averaging periods and regions, transport model, and mission concepts.

This study has quantified the benefit of A-SCOPE in conjunction with the ground-based flask sampling network only. Both observational types are similar in that they constrain the net flux of carbon dioxide via the atmospheric concentration. There are or will be, however, further remote sensing data streams available from optical sensors (e.g.

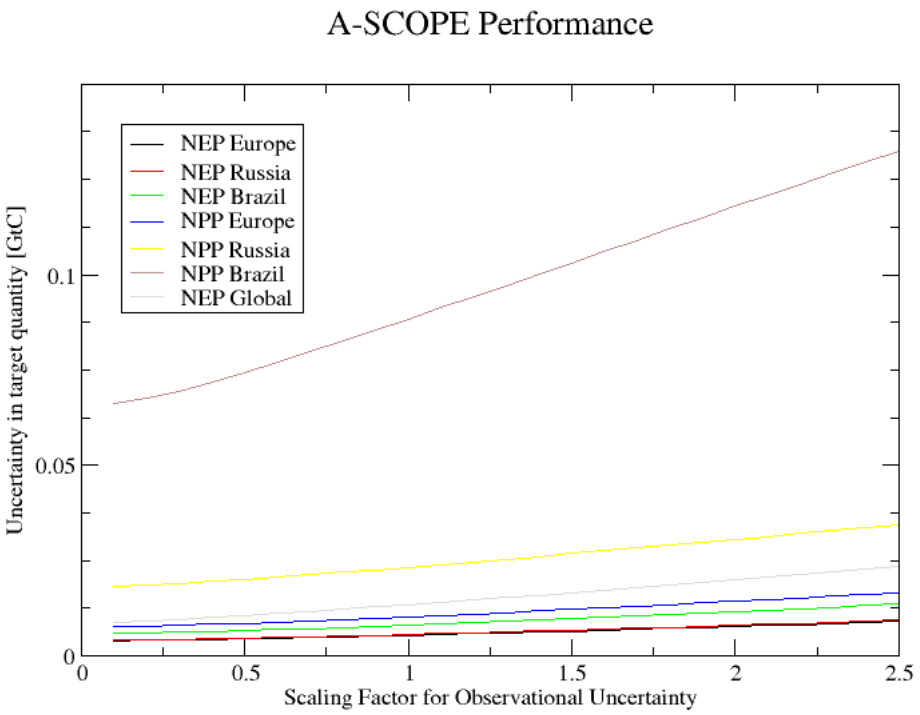


**Figure 5.** Posterior uncertainty for A-SCOPE only as a function of observational uncertainty.

MERIS) or microwave sensors (e.g. SMOS). Such instruments provide direct constraints on vegetation phenology (MERIS) or hydrology (SMOS), which are tightly coupled to the terrestrial carbon cycle. It is expected that such data streams provide constraints complementary to A-SCOPE. It is highly desirable to set up a system that can benefit of this multiple constraint in terms of uncertainty reduction in carbon fluxes. Since standard transport inversions lack any process representation of terrestrial phenology and hydrology, such a quantitative assessment has to be performed in the framework of a CCDAS.

## 5 Acknowledgments

This study was supported by the European Space Agency under contract 21061/07/NL/JA. The authors thank Simon Blessing for his help with shell-scripting, Martin Heimann for providing the atmospheric transport model TM3, and Christiane Textor for helpful comments on the manuscript. Computing support was provided by the QUEST computer cluster at the University of Bristol, UK.

## REFERENCES

Baker, D. F., Bösch, H., Doney, S. C. and Schimel, D. S. 2008. Carbon source/sink information provided by column $CO_2$ measurements from the Orbiting Carbon Observatory. *Atmospheric Chemistry and Physics Discussions* **8**(6), 20051–20112.

Breon, F. M., Houweling, S., Aben, I., Ehret, G., Chevallier, F., Marshall, J., Flamant, P., Bruneau, D. and Rayner, P. 2009. Observation Techniques and mission concepts for the Analysis of the Global Carbon Cycle, ESA contract 20839/07/NL/HE, *Technical report*, European Space Agency, Noordwijk, The Netherlands.

Cadule, P., Friedlingstein, P., Bopp, L., Sitch, S., Jones, C. D., Ciais, P., Piao, S. L. and Peylin, P. 2010. Benchmarking coupled climate-carbon models against long-term atmospheric $CO_2$ measurements. *Glob. Biogeochem. Cyc., in press* .

Chevallier, F., Breon, F.-M. and Rayner, P. J. 2007. Contribution of the Orbiting Carbon Observatory to the estimation of $CO_2$ sources and sinks: Theoretical study in a variational data assimilation framework. *J. Geophys. Res.* **112**, doi:10.1029/2006JD007375.

Chevallier, F., Maksyutov, S., Bousquet, P., Breon, F.-M., Saito, R., Yoshida, Y. and Yokota, T. 2009. On the accuracy of the $CO_2$ surface fluxes to be estimated from the GOSAT observations. *Geophys. Res. Lett,* **36**. doi:10.1029/2009GL040108.

Cramer, W., Bondeau, A., Woodward, F., Prentice, I., Betts, R. et al. 2001. Global response of terrestrial ecosystem structure and function to $CO_2$ and climate change: results from six dynamic global vegetation models. *Global Change Biology* **7**(4), 357–373.

Cramer, W., Kicklighter, D. W., Bondeau, A., Moore, B. I., Churkina, G., Ruimy, A., Schloss, A. L., Kaduk, J. and the participants of the Potsdam NPP Model Intercomparison 1999. Comparing global models of terrestrial net primary productivity (NPP): overview and key results. *Glob. Change Biol.* **5**, 1–15.

Ehret, G., Kiemle, C., Wirth, M., Amediek, A., Fix, A. and Houweling, S. 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 Enting, I. G. Cambridge University Press.

ESA 2008. A-SCOPE - Advanced Space Carbon And Climate Observation of Planet Earth, *Technical Report SP-1313/1*, European Space Agency, Noordwijk, The Netherlands.

Feng, L., Palmer, P. I., Bsch, H. and Dance, S. 2009. Estimating surface co$_2$ fluxes from space-borne co$_2$ dry air mole fraction observations using an ensemble kalman filter. *Atmospheric Chemistry and Physics* **9**(8), 2619–2633.

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

Giering, R. and Kaminski, T. 1998. Recipes for Adjoint Code Construction. *ACM Trans. Math. Software* **24**(4), 437–474.

GLOBALVIEW-$CO_2$ 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].

Heimann, M. and Kaminski, T.: 1999, Inverse modeling approaches to infer surface trace gas fluxes from observed atmospheric mixing ratios, *in* A. F. Bouwman (ed.), *Approaches to scaling of trace gas fluxes in ecosystems*, Elsevier, Amsterdam, chapter 14, pp. 275–295.

Heimann, M. and Körner, S. 2003. The global atmospheric tracer model TM3, *Technical Report 5*, Max-Planck-Institut für Biogeochemie, Jena, Germany.

Houweling, S., Breon, F.-M., Aben, I., Rödenbeck, C., Gloor, M., Heimann, M. and Ciais, P. 2004. Inverse modeling of $CO_2$ sources and sinks using satellite data: a synthetic intercomparison of measurement techniques and their performance as a function of space and time. *Atmos. Chem. Phys.* **4**, 523–538.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D. et al. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* **77**(3), 437–471.

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

Kaminski, T. and Heimann, M. 2001. Inverse modeling of atmospheric carbon dioxide fluxes. *Science* **294**(5541), 259.

Kaminski, T., Knorr, W., Rayner, P. and Heimann, M. 2002. Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Glob. Biogeochem. Cyc.* **16**(4), doi:10.1029/2001GB001463.

Kaminski, T. and Rayner, P. J.: 2008, Assimilation and network design, *in* H. Dolman, A. Freibauer and R. Valentini (eds), *Observing the continental scale Greenhouse Gas Balance of Europe*, Ecological Studies, Springer-Verlag, New York, chapter 3, pp. 33–52. doi:10.1007/978-0-387-76570-9_3.

Keeling, C. and Whorf, T. 2002. Atmospheric $CO_2$ records from sites in the SIO air sampling network. *Trends: A Compendium of Data on Global Change* pp. 723–741.

Kicklighter, D. W., Bondeau, A., Schloss, A. L., Kaduk, J., McGuire, A. D. and the participants of the Potsdam NPP Model Intercomparison 1999. Comparing global models of terrestrial net primary productivity (NPP): global pattern and differentiation by major biomes. *Glob. Change Biol.* **5**, 16–24.

Knorr, W.: 1997, Satellitengestützte Fernerkundung und Modellierung des Globalen $CO_2$ -Austauschs der Landvegetation: Eine Synthese, PhD thesis, Max-Planck-Institut für Meteorologie, Hamburg, Germany.

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

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

Michalak, A., Miller, C., Browell, E., Moore, B., Abshire, J., Spiers, G., Jucks, K. and Emanuel, B. 2008. ASCENDS – Active Sensing of $CO_2$ Emissions over Nights, Days, and Seasons Mission, *Technical report*, University of Michigan in Ann Arbor, Michigan.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls, M. E., Doney, S. C., Pawson, S., Boesch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L. and Law, R. M. 2007. Precision requirements for space-based $XCO_2$ data. *J. Geophys. Res.* **112**, doi:10.1029/2006JD007659.

Pak, B. C. and Prather, M. J. 2001. $CO_2$ source inversions using satellite observations of the upper troposphere. *Geophys. Res. Lett.* **28**, 4571–4574.

Patra, P. K., Sasano, S. M. Y., Nakajima, H. and Inoue, G. 2003. An evaluation of $CO_2$ observations with SOFIS sensor for surface source inversion. *J. Geophys. Res.* **108**, doi:10.1029/2003JD003661.

Randerson, J., Hoffman, F., Thornton, P., Mahowald, N., Lindsay, K., Lee, Y.-H., Nevison, C., Doney, S., Stöckli, R., Bonan, G., Covey, C., Running, S., and Fung, I. 2009. Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. *Glob. Change Biol.* **15**, 2462–2484.

Rayner, P. J., Enting, I. G. and Trudinger, C. M. 1996. Optimizing the $CO_2$ observing network for constraining sources and sinks. *Tellus* **48B**, 433–444.

Rayner, P. J., Law, R. M., O'Brien, D. M., Butler, T. M. and Dilley, A. C. 2002. Global observations of the carbon budget: 3. Initial assessment of the impact of satellite orbit, scan geometry and cloud on measuring $CO_2$ from space. *J. Geophys. Res.* **107**(D21), 4557. doi:10.1029/2001JD000618.

Rayner, P. J. and O'Brien, D. M. 2001. The utility of remotely sensed $CO_2$ concentration data in surface source inversions. *Geophys. Res. Let.* **28**, 175–178.

Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and Widmann, H. 2005. Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS). *Global Biogeochemical Cycles* **19**, doi:10.1029/2004GB002254.

Scholze, M.: 2003, Model studies on the response of the terrestrial carbon cycle on climate change and variability, PhD thesis, Max-Planck-Institut für Meteorologie, Hamburg, Germany.

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

Tarantola, A. 1987. Inverse problem theory - methods for data fitting and model parameter estimation Tarantola, A. Elsevier Sci., New York.