Biogeosciences Discuss., 9, 3615–3643, 2012
www.biogeosciences-discuss.net/9/3615/2012/
doi:10.5194/bgd-9-3615-2012

Biogeosciences Discussions

Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana

T. Kato[1,2,3], M. Scholze[1,4], W. Knorr[1,5], E. Veenendaal[6], T. Kaminski[7], J. Kattge[8], and N. Gobron[9]

[1]Department of Earth Sciences, University of Bristol, Bristol, UK
[2]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[3]Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ, CEA-orme des Merisiers, 91191 Gif-sur-Yvette, France
[4]KlimaCampus, University of Hamburg, Hamburg, Germany
[5]Department of Meteorology and Climatology, Aristotelian University of Thessaloniki, Thessaloniki, Greece
[6]Nature Conservation and Plant Ecology Group, Department of Environmental Sciences,Wageningen University, Wageningen, The Netherlands
[7]FastOpt, Hamburg, Germany
[8]Max-Planck-Institute for Biogeochemistry, Jena, Germany

[9]European Commission, Joint Research Center, Ispra, Italy

Received: 7 March 2012 – Accepted: 12 March 2012 – Published: 22 March 2012

Correspondence to: T. Kato (tomomichi.kato@lsce.ipsl.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

BGD
9, 3615–3643, 2012

Abstract

Terrestrial productivity in semi-arid woodlands is strongly susceptible to changes in precipitation, and semi-arid woodlands constitute an important element of the global water and carbon cycles. Here, we use the Carbon Cycle Data Assimilation System (CCDAS) to investigate the mechanisms controlling ecological and hydrogical activities for a semi-arid savanna woodland site in Maun, Botswana. Twenty-four eco-hydrological process parameters of a terrestrial ecosystem model are optimized against two data streams either separately or simultaneously: daily averaged latent heat flux (LHF) derived from eddy covariance measurement, and decadal fraction of absorbed photosynthetically active radiation (FAPAR) derived from Sea-viewing Wide Field-of-view Sensor (SeaWiFS).

Assimilation of both LHF and FAPAR for the years 2000 and 2001 leads to improved agreement between measured and simulated quantities not only for LHF and FAPAR, but also for photosynthetic CO_2 uptake. The closest agreement is found for each observed data stream when only the same data stream is assimilated. The mean uncertainty reduction (relative to the prior) over all parameters is 16.1 % for the simultaneous assimilation of LHF and FAPAR, 9.2 % for assimilating LHF only, and 7.8 % for assimilating FAPAR only. Furthermore, the set of parameters with the highest uncertainty reduction is similar between assimilating only FAPAR or only LHF. The highest uncertainty reduction is found for a parameter describing maximum plant-available soil moisture for all three cases. This indicates that not only LHF but also satellite-derived FAPAR data can be used to constrain and indirectly observe hydrological quantities.

1 Introduction

Terrestrial ecosystems are strongly interconnected with the climate system through the hydrological cycle by various processes, such as infiltration, runoff, evaporation and transpiration. In particular, latent heat flux (LHF), resulting from the sum of evaporation

BGD
9, 3615–3643, 2012

and transpiration, is an essential component of the surface energy balance and needed for understanding the global and local water balance. It is also a key quantity for understanding the physiological response of ecosystems to changes in climate, as LHF is related to the terrestrial carbon cycle through stomatal function and leaf size. Information on the latent heat fluxes of terrestrial ecosystems can improve our understanding of ecosystem functioning and its potential response to changes in the Earth's climate through anthropogenic interference, such as an increased frequency of droughts (IPCC, 2007).

To fill the gap between measurements of terrestrial ecosystem fluxes and eco-physiological theory as embodied in terrestrial ecosystem models, data assimilation techniques are becoming more widely used in biogeochemistry. The main application of such data assimilation systems is focussed on the optimisation of model process parameters, primarily against observations of the carbon cycle, e.g. atmospheric CO_2 concentration, carbon fluxes and pools (e.g. Rayner et al., 2005; Braswell et al., 2005; Williams et al., 2005; Knorr and Kattge, 2005). These studies provide, besides parameters optimized to fit model output to observations, a better understanding of the key processes controlling the ecosystem behaviour with regard to eco-physiological functioning and, closely related, ecosystem carbon cycling.

Here we use the fully variational Carbon Cycle Data Assimilation System (CCDAS). It has been designed to estimate process parameters through assimilation against observations, mainly atmospheric CO_2 concentration from ground-based measurement stations and fraction of absorbed photosynthesically active radiation (FAPAR) from satellite on a global scale (Rayner et al., 2005; Scholze et al., 2007; Kaminski et al., 2011). CCDAS is based on a variational approach and makes use of the availability of the adjoint (1st derivative) model to optimize parameters. Furthermore, CCDAS is able to calculate posterior parameter uncertainties through use of the Hessian matrix (2nd derivative of the misfit function between model and data) and propagate these uncertainties through the model to diagnostic and prognostic quantities of interest, e.g. the

BGD
9, 3615–3643, 2012

net carbon flux. CCDAS has so far been applied for assimilation of atmospheric CO_2 concentration and FAPAR observations.

In this study, CCDAS is extended to assimilate LHF and to estimate further parameters related to the hydrological part of the model. LHF is calculated in conjunction with terrestrial carbon fluxes by the land surface model BETHY (Biosophere Energy-Transfer Hydrology; Knorr 2000) and match, within the assimilation scheme, to LHF measured with eddy covariance (EC) systems. The aim of this work is thus to broaden the capability of CCDAS, and make it applicable at further temporal and spatial scales, and opening up its application to the study of eco-hydrological processes.

Savannas are climatically characterized by a distinct seasonality of rainfall, i.e. a combination of a severe dry season and a moderate wet season. Therefore, Savanna vegetation is adapted to dry conditions and usually composed of sparse trees and grasses, whose canopy does not close. These regions are potentially at risk from large changes in the seasonality of water availability as well as the total amount of available water caused by climate change. For example Wang (2005) showed that the models consistently predicted less rainfall and consequently drier soils at the end of the 21st century over much of subtropical and temperate regions including savannas.

Recent model studies have analysed the importance of various processes on the hydrological conditions in savanna ecosystems. For example, Kleidon and Heimann (1996) and later Ichii et al. (2009) highlighted the importance of rooting depth within land surface models, which deal with both LHF and carbon fluxes, assuming that ecosystems are maximizing their productivity under water-limited conditions.

To investigate eco-hydrological dynamics of ecosystems as a whole, the eddy covariance (EC) technology has been applied in various terrestrial ecosystems (Aubinet et al., 2000; Baldocchi et al., 2001) as a reliable way of measuring the in-situ energy, water and carbon fluxes. Compared to closed-forest ecosystems in the Northern Hemisphere, however, little attention has been given to savanna ecosystems, even though they cover approximately $17 \times 10^6\,km^2$ globally (around 20 % of terrestrial surface), more than either temperate or boreal forests (Veenendaal et al., 2004). Recent efforts

BGD
9, 3615–3643, 2012

in conducting eddy covariance observations in savanna regions (Veenendaal et al., 2004) enable us to greatly improve our modelling capabilities, and to better understand eco-hydrological functioning in open canopy woodlands.

In this study, we extend the original CCDAS to be able to assimilate eddy-covariance measurements of LHF and optimise model process parameters included in all components (energy, carbon and water balance, phenology) of the BETHY model. We apply CCDAS to simultaneously assimilate eddy-covariance LHF and remotely-sensed FAPAR observations at a single point for a semi-arid savanna site at Maun, Botswana. To our knowledge, this is so far the first attempt to consistently assimilate these two different data streams into a terrestrial ecosystem model using the adjoint-based gradient approach.

2 Materials and methods

2.1 Site description and measurement data

We have selected a Mopane tree woodland area at Maun, Botswana (23°33′ E, 19°54′ S; 950 m a.s.l.; Veenendaal et al., 2004). With a canopy cover of 30–40 %, the plant community at the flux measurement site is dominated by the mopane tree (*Colophospermum mopane*), and the marginal under-story consists of grasses with a canopy cover of at most 15 %, dominated by *Panicum maximum*, *Schmidtia pappophoroides* and *Urochloa trichopus*. The mean maximum and minimum temperatures of the warmest and coldest month and annual precipitation are 33.6 and 7.1 °C and 464 mm, respectively. There is a distinct dry season during the winter months from May to September. Substantial amounts of rainfall are normally limited to between December and March.

LHF and CO_2 flux measurements are conducted by the EC method using a 12.6 m high tower in the middle of a homogeneous tall mopane tree stand with a maximum canopy height of about 8 m (Veenendaal et al., 2004). Three-dimensional wind speed,

BGD
9, 3615–3643, 2012

humidity, and CO_2 concentration were logged with a frequency of 20 Hz, and fluxes are integrated into half hour means with the EdiSol software (Moncrieff et al., 1997). Moreover, air temperature, shortwave radiation, and precipitation are also measured at the same tower, and are used to calibrate the climate input data, which is extracted from a global data set as described in the following paragraph. In this study, data from 2000 and 2001 are used for assimilation. Missing data due to unfavourable meteorological or instrumental conditions, have been replaced by a gap-filling scheme (see Appendix A of this paper). Due to missing half-hourly data, only 223 points of daily averaged LHF data out of 731 days for two years would be available for assimilation if we restricted ourselves to complete diurnal measurement cycles. To get both a sufficient number of data points and avoid biases in daily averaged LHF values from the gap-filling procedure, we include data points where up to four points of 48 half-hourly values were gap-filled within one day (see Appendix A for the gap-filling scheme used). This yields a total of 464 daily data points for the two selected years 2000 and 2001.

Input data of daily precipitation, daily minimum and maximum temperatures and incoming solar radiation at the site are derived from a global gridded climate data set, generated through a combination of available monthly gridded and daily station data (R. Schnur, personal communication, 2008) by a method by Nijssen et al. (2001), using gridded data from the Summary of the Day Observations (Global CEAS), National Climatic Data Center and the latest updates of gridded data by Jones et al. (2001) and Chen et al. (2002). These data are then corrected using the local climatology measured at the eddy flux tower (Lloyd et al., 2004). This is done by deriving linear regression equations between daily minimum and maximum temperatures and incoming solar radiation from the global data set and the local measurements. Daily precipitation from the global data set is adjusted by multiplying the global data with a constant factor such that the total rainfall matches that of the local rainfall data.

The assimilated FAPAR observations are derived from the Sea-viewing Wide Field-of-view (SeaWiFS) of the National Aeronautics and Space Administration NASA at a spatial resolution of 1.5 km (Gobron et al., 2006). The FAPAR data are provided every

BGD
9, 3615–3643, 2012

10 days as representative values over the period giving a total of 70 data points over the two-year study period. 3 by 3 pixel scenes centred around the position of the Maun flux site are used here.

We have chosen the Maun site for several reasons. First, two years of flux data, both LHF and carbon fluxes, measured by the EC technique during the SAFARI2000 campaign are available (Lloyd et al., 2004). Second, a flat topography and homogeneous land cover increase the accuracy of EC data and also of the FAPAR satellite observations. There are also significantly fewer cases of cloudy conditions at a savanna site as compared to e.g. tropical forest sites. Third, the dominant land cover types with Mopane trees, understory grasses, and patchy bare ground, which change their relative coverages seasonally, are potentially responsible for large amplitudes and distinct seasonality in LHF and other related quantities. This environment thus provides a welcome opportunity for testing and enhancing the capability of CCDAS (improving model formulation and parameter settings) in an area with water-limited conditions and low productivity.

2.2 Carbon Cycle Data Assimilation System

In its original version, CCDAS combines the land biosphere model BETHY (Knorr, 2000) with the atmospheric tracer transport model TM2 (Heimann, 1995) and some background fluxes not computed by BETHY (fossil fuel and land use change emissions and ocean-atmosphere exchange fluxes) to simulate the terrestrial carbon cycle globally along with atmospheric CO_2 concentrations. It uses first and second derivatives to optimize internal model process parameters and subsequently derive posterior uncertainties on these parameters. In this study, we modify the version of CCDAS as described in Knorr et al. (2010), to assimilate LHF and FAPAR at the Maun savanna site. BETHY calculates ecosystem energy, water and CO_2 fluxes using the before-mentioned climate input data. Two plant functional types (PFTs), namely tropical broadleaf deciduous tree with a warm-deciduous phenology and C4 grass (recognized as PFT 2 and 10, respectively, in the original BETHY model), are simulated

BGD
9, 3615–3643, 2012

for the Maun site with a fractional coverage of 0.7 and 0.3 for PFT 2 and 10, respectivly. Detailed information about BETHY is given in the Appendix B in the Supplement. Differences between simulated LHF and FAPAR values and the observed data are minimized by optimizing model process parameters. Here we only briefly summarise the main methodological aspects. For detailed information on the CCDAS methodology we refer to Scholze et al. (2003), Rayner et al. (2005), and Scholze et al. (2007).

2.3 Cost function and observational uncertainties

The cost function $J(\boldsymbol{p})$ ($\boldsymbol{p}$ denotes the parameter vector) expresses the differences between simulated and observed quantities normalised by the uncertainty of each of the contributing observations, LHF and FAPAR, under the assumption of a Gaussian probability density distribution. It is formulated in a Bayesian form:

$$J(\boldsymbol{p}) = \frac{1}{2}\left[\boldsymbol{p}-\boldsymbol{p}_0\right]^T \boldsymbol{C}_{p_0}^{-1}\left[\boldsymbol{p}-\boldsymbol{p}_0\right] + \frac{1}{2}\left[\boldsymbol{e}(\boldsymbol{p})-\boldsymbol{e}_0\right]^T \boldsymbol{C}_{e_0}^{-1}\left[\boldsymbol{e}(\boldsymbol{p})-\boldsymbol{e}_0\right] + \frac{1}{2}\left[\boldsymbol{a}(\boldsymbol{p})-\boldsymbol{a}_0\right]^T \boldsymbol{C}_{a_0}^{-1}\left[\boldsymbol{a}(\boldsymbol{p})-\boldsymbol{a}_0\right] \quad (1)$$

Where $\boldsymbol{p}$ is the parameter vector, $\boldsymbol{p}_0$ is the prior parameter vector (0 denotes the prior value), C_{p0} the uncertainty for the prior parameter vector $\boldsymbol{p}_0$ in the form of a covariance matrix. $\boldsymbol{e}(\boldsymbol{p})$ and $\boldsymbol{a}(\boldsymbol{p})$ are modeled LHF and FAPAR values as a function of the parameter set $\boldsymbol{p}$, $\boldsymbol{e}_0$ and $\boldsymbol{a}_0$ are the observations of LHF and FAPAR, and C_{e0} and C_{a0} express the uncertainties of the observations $\boldsymbol{e}_0$ and $\boldsymbol{a}_0$. ^{T} and $^{-1}$ denote the transpose and inverse of matrices. J is minimized iteratively using derivative information calculated by the adjoint model.

The Hessian matrix, the second derivative of J with respect to the parameters, is used to estimate posterior parameter uncertainties, using the mathematical property that the inverse of the Hessian matrix at the cost function minimum approximates the posterior parameter error covariance matrix. All derivative code is derived efficiently from the models' source code (see Kaminiski et al., 2003) by applying the automatic

BGD
9, 3615–3643, 2012

differentiation tool TAF (Transformation of Algorithms in Fortran; Giering and Kaminski 1998).

The uncertainty of observational LHF is taken as the higher of either 10.0 W m^{-2} or 23 % of measured LHF. The 23 % threshold was derived from the energy imbalance at this site, which was calculated as the underestimation of the sum of daily averaged sensible and latent heat flux (SHF + LHF) compared to the sum of net radiation, Rn, and soil heat flux, G, (Rn + G) in the regression line: SHF + LHF = 0.77 (Rn + G) – 12.2 W m^{-2}, r^2 = 0.79. As in Knorr et al. (2010) the uncertainty of observational FAPAR is set to a constant value of 0.1 for all observations.

2.4 Eco-hydrological parameters

We select 24 parameters to be optimized against observed LHF and FAPAR data (Table 2). 18 of the 24 parameters are related to the model's physiology and phenology, and have been optimized with CCDAS in previous works (e.g. Knorr et al., 2010). The prior mean and uncertainty values for these 18 parameters are the same as those used in previous studies (Scholze et al., 2007; Knorr et al., 2010). The six new parameters describe the water balance or the interaction between water and carbon fluxes. They are $f\mathrm{ci}_{\mathrm{C3}}$ and $f\mathrm{ci}_{\mathrm{C4}}$, the ratio of CO_2 concentration inside and outside leaf tissues for C3 and C4, respectively (Eq. (A21) in the Supplement); C_{W}, the ratio of maximum water supply rate from the roots relative to plant available soil moisture (Eq. (A24) in the Supplement); h_0, a scaling factor of the relative dryness of air (Eq. (A35) in the Supplement); $\hat{h}$, a scaling factor of the relative humidity of air (Eq. (A34) in the Supplement); and W_{max}, the maximum plant-available soil moisture (Eq. (A24) in the Supplement). The former five parameters have been first introduced in Knorr and Heimann (2001) and further information is given in the Supplement. The parameter W_{max} scales the maximum amount of plant-available soil moisture. A decreasing value of this parameter is reflected in a decline of relative soil wetness, leading to less evapotranspiration, in BETHY. These six parameters control stomatal aperture, energy balance, and water

BGD
9, 3615–3643, 2012

balance processes in BETHY. They are important for eco-hydrological functioning under semi-arid conditions in savanna ecosystems.

2.5 Experimental set-up

To investigate the impact of multiple data streams on the assimilation results, we perform three assimilation experiments: (1) assimilating only LHF data, (2) assimilating only FAPAR data, and (3) assimilating LHF and FAPAR data simultaneously. Prior/posterior simulations with only LHF, with only FAPAR, and with a combination of both data will be noted as prior/posterior experiment 1, 2, and 3, hereafter. The assimilation experiment with only LHF data considers the first and second terms in Eq. (1), that with only FAPAR considers the first and third terms, and that with both types of data combined considers all three terms.

3 Results

3.1 Optimization and parameter uncertainty

The optimization takes 27, 29, and 41 iterations to converge to a minimum for experiments 1, 2 and 3, respectively. The total value of the cost function could be reduced substantially (see Table 2). Also, the gradient of the cost function was reduced from 82, 93, 170 for experiments 1, 2 and 3, respectively, to a final value close to zero (of the order of 10^{-2} to 10^{-7}, see Table 1).

Table 2 and Fig. 5 show the posterior parameter values and their uncertainty reduction relative to the prior (defined as $1 - \sigma_{\text{posterior}}/\sigma_{\text{prior}}$) for all three experiments along with their prior values and uncertainties. For all experiments, the parameter $W_{\max}$ decreases substantially from the prior value of 1500 mm, to 332 mm, 86 mm, and 129 mm, for experiments 1, 2, and 3, respectively.

Relative parameter uncertainties are reduced by more than 20 % for three, four, and six of the 24 parameters for experiments 1, 2, and 3, respectively. Note that the largest

BGD
9, 3615–3643, 2012

uncertainty reductions always occur in experiment 3 where we simultaneously assimilate both data streams. The parameters with high uncertainty reductions differ slightly among the experiments (Table 2 and Fig. 5). In general, parameters showing considerable uncertainty reductions are: the maximum catalytic capacity of rubisco (V_{max}^{25}; parameters 1 and 2), the expected length of drought periods tolerated before leaf shedding (τ_W; parameters 17 and 18), the standard ratio of CO_2 concentration inside and outside the leaf tissus for C3 plant ($f ci_{C3}$; parameter 19), and the maximum plant-available soil moisture (W_{max}: parameter 24). Also, the change in parameter values appears to be large for the above six parameters (Table 2). On the other hand, some of the remaining 18 parameters, for which the uncertainty reduction is less than 20 %, also show relatively large deviations (in relation to their prior uncertainty) from their prior parameter value (Table 2). More specifically, they are related to the activation energies and Michaelis-Menten constants of the temperature dependency of further enzyme kinetics, E_{Vmax} and K_C^{25}, as well as the efficiency of electron transport, α_q for C3, and the linear growth constant in LAI , ξ (see Supplement for a more specific explanation of the parameters). For the other 14 parameters, both the posterior value and the uncertainty hardly change compared to the respective prior values.

3.2 LHF and FAPAR

Compared to the measurements, the simulated prior LHF values have a too small seasonal amplitude with lower values during the wet season (except for some scattered points between November to April) and slightly higher values than the observations during the dry period (Fig. 2). Experiments 1 and 3 show a reasonable seasonality of LHF with high values in the wet season gradually declining during the dry season, starting from April and ending in October, with slightly lower values than the observations. However, experiment 2 gives lower LHF values than the observations over almost the entire simulation period although with some scattered high values in the wet season. This results in the highest root mean square error (RMSE) of 26.5 $W m^{-2}$ compared to RMSE values of 14.6, and 21.5 $W m^{-2}$ for experiments 1 and 3, respectively.

BGD
9, 3615–3643, 2012

Prior FAPAR values are nearly constant around a high value of 0.94 (averaged prior value over the 2 yr, see Fig. 3), however, the observations show much lower values with a distinct seasonality ranging between 0.11 and 0.39. In experiment 1, the modeled FAPAR values during the wet season are similar to those from the prior run, with high, nearly constant values, but with the difference that they decrease to about 0.6 in the dry season. Modelled FAPAR values in experiment 2 show a good agreement with the observations indicated by a small RMSE of 0.06. Here, the simulated FAPAR values fall within the uncertainty range of the observed values over almost the entire period. In experiment 3, the modeled FAPAR values have a distinct seasonality with values larger than the observations during the wet period, but showing a good agreement with the observations at the end of the dry period in October. The RMSE for experiment 3 is as low as 0.20, while those for the prior simulation and experiment 1 are 0.73 and 0.68, respectively.

4 Discussion

4.1 Constraint of parameters by eddy water flux and satellite FAPAR data

W_{max} is consistently constrained to a relatively small value in all three experiments (86 to 332 mm) in contrast to the general belief that rooting depth would be large in such dry conditions, leading to a large value for the maximum plant-available soil water. For example, Schenk and Jackson (2002) suggested that dry tropical savannas have on average a rooting depth of 1.44 m containing 95 % of the total ecosystem roots. In fact, Veenendaal et al. (2008) showed that the tall and short mopane trees rooted at least deeper than 1.0 m by field measurement in the field measurement at Maun site, However, they also indicated that total root density of both type mopanes as well as fine root dentisity of short mopane were concentrated in the upper soil fraction up to 20 cm depth. That suggests that active layer for soil water uptake would be in shallow soil layer, partly supporting the small W_{max} in our simulations.

BGD
9, 3615–3643, 2012

The posterior error covariance matrixes of the 24 parameters show values of less than 0.1 for the error covariances of W_{max} with all other parameters (Figs. A1, A2, A3), suggesting that W_{max} can be independently constrained by LHF and FAPAR observations. It is also surprising that all three experiments yield essentially the same result for W_{max}. This also means that by assimilating FAPAR observations only, we can constrain maximum plant available soil moisture, a key parameter of the soil water balance, and thus quantify the hydrological state of the ecosystem. The slightly larger W_{max} value in experiment 1 simply relates to the larger simulated LAI values, which require larger transpiration rates driven by enhanced root water uptake that leads to a large rooting depth.

Although the optimized value of Λ_{max} (parameter for the environmental limitation of maximum LAI) does not differ among three experiments, the modeled maximum leaf area index (LAI) differs largely. Compared to the observed annual maximum in LAI between 0.9 and 1.3 (Mantlana, 2002; Veenendaal et al., 2008), experiment 2 and 3 show relatively reasonable simulated annual maximum LAI values of 0.99 and 2.92, respectively. However, experiment 1 gives a value of 5.20, which is close to the environmental limitation given by the parameter value of 5.22 for Λ_{max} (Table 2). In addition, the modelled LAI shows only a small seasonal amplitude similar to modelled FAPAR in experiment 1 (Fig. 3). Therefore assimilating only LHF data leads to too high LAI values with a poor seasonality, although the simulated LHF fit the observations very well in experiment 1. We assume that the small τ_W value for PFT 2 of 10 days could be a potential cause for the large and constant LAI value. In contrast to experiment 1, τ_W shows much higher posterior values of 186 days and 94 days in experiments 2 and 3, which possibly leads to a larger seasonality and also a lower LAI. Recall that τ_W represents the expected length of drought periods tolerated before leaf shedding (Knorr et al., 2010). Such a small τ_W value reflects the plant's water reserves to be sufficient for continued plant growth, and thus induces growth of LAI when there is a certain level of soil water. The high reduction in the relative parameter uncertainty of

BGD
9, 3615–3643, 2012

τ_W for PFT 2 and 10 by more than 30 % suggests a strong constraint by the FAPAR observations on the phenology component of BETHY, as expected.

The relative uncertainty reduction for parameter V_{max}^{25} for C3 (42 %) as well as for parameters ξ (14 %) are substantially larger in experiment 3 than in experiments 1 and 2 with relative uncertainty reductions for these parameters between 0 % (ξ) and 3 % (V_{max}^{25} and ξ, see Table 2 and Fig. 5). This suggests that each data stream carries complementary information on photosynthesis and phenology such that the combined assimilation has the apparent strong constraint on specific paramters of plant productivity and leaf phenology. Interestingly, in experiment 3 the V_{max}^{25} parameter show fairly high negative error covariances with the respective $f$ci parameter (ratio of CO_2 concentration inside the leaf tissue to the outside concentration) of −0.43 for C3 trees and −0.25 for C4 grass as shown in Table A3 in the Supplement. This together with the increased $f\mathrm{ci}_{C3}$ value explains at least to some extent the small posterior V_{max}^{25} value of 34 for PFT 2 in experiment 3 because a higher $f\mathrm{ci}_{C3}$ value increases the CO_2 uptake of plants by photosynthesis.

It is interesting to note that posterior magnitudes of the absolute values in LHF, FAPAR and also the photosynthetic CO_2 uptake flux (Gross Primary Production, GPP, see also next section) show a somewhat intermediate value in experiment 3 when assimilating both data streams simultaneously as compared to their values in experiments 1 and 2 when assimilating only one of the two data streams at a time as can be seen in Figs. 2–4: experiment 1 yields the largest values, experiment 2 the smallest and experiment 3 somewhat intermediate values. This suggests that the posterior parameter values are a compromise as the optimization is not capable of simultaneously fitting each of the two data stream as good as in the experiments 1 and 2 in which only one of the two data streams is assimilated.

4.2 Simulation of carbon fluxes

Figure 4 displays the Gross Primary Production (GPP), which is the total amount of CO_2 that is taken up by plants by photosynthesis, simulated by BETHY. Experiment

BGD
9, 3615–3643, 2012

1 gives the highest GPP among the three experiments, which is mainly an effect of the simulated large LAI values. Experiment 2 shows much lower simulated GPP than the observations and the simulated GPP is in fact the lowest among the three experiments. However, FAPAR is quite well reproduced compared to the satellite data (Fig. 3) suggesting that simulated GPP should be much closer to the observations. This discrepancy leads us to the assumption that the FAPAR observations are possibly biased towards lower values than true ground data would suggest due to cloud contamination during the wet season (see also below).

Besides a good fit of LHF in experiment 3, the simulated GPP also shows a moderately good fit in seasonality, which is reflected in a lower RMSE than in experiment 2. This somehow balanced result in terms of fitting independent observations such as the gross carbon flux in experiment 3 is similar to results from Barbu et al. (2011). They showed that when assimilating LAI observations in addition to assimilating soil wetness index observations, the root mean square error of simulated net ecosystem CO_2 fluxes with observed fluxes was reduced by about 5 % for a grassland site in southwest France. We suppose that the reduced RMSE for GPP in our case for experiment 3 is caused by changes in multiple parameters: larger $f\mathrm{ci}_{C3}$ and $f\mathrm{ci}_{C4}$ than in experiment 1 and 2, and an increased value for τ_W for PFT 2 as compared to the prior value (Table 2). Larger $f\mathrm{ci}_{C3}$ and $f\mathrm{ci}_{C4}$ increase the transpiration rate and also the carbon exchange per unit leaf area by increasing stomatal aperture. Larger τ_W, as explained above, adjusts the seasonality in FAPAR and LAI, and partly also the seasonality in GPP.

4.3 Simultaneous assimilation of multiple data sets

Experiments 1 and 2 show that we are able to adequately simulate the data stream (LHF in experiment 1 and FAPAR in experiment 2) separately but at the expense of an inferior fit to the respective other data stream. On the other hand, the simultaneous assimilation of LHF and FAPAR observations in experiment 3 indicates a good agreement with both observations. It is noteworthy to point out that FAPAR observations contribute

BGD
9, 3615–3643, 2012

to the reduction of parameter uncertainty of the hydrological parameter, V_{max}^{25} for C3 even in experiment 3, demonstrated by the fact that uncertainty reduction for these parameters is higher in experiment 3 than in experiment 1. FAPAR alone also helps to constrain W_{max} in experiment 2. This finding that FAPAR is able to constrain hydrological parameters is confirmed by Kaminski et al. (2011). They showed that the assimilation of FAPAR in CCDAS was able to constrain not only hydrological parameters on a global scale, but also estimates of soil moisture and evapotranspiration, alone or in combination with atmospheric CO_2 concentrations.

Comparing the results from all three experiments reveals the potential problem in FAPAR. In experiment 3 both seasonality and magnitude of LHF are simulated adequately (similar to the results in experiment 1) but here the model is not able to simulate the low FAPAR observations during the wet season. However, experiment 2 shows a very good fit to the FAPAR observations over the entire period, but the model underestimates the amplitude in LHF during the wet season (Fig. 2). From these results we believe that the mismatch between simulated and observed FAPAR in experiment 3 during the wet period is caused by a bias towards lower values in the FAPAR observations during the wet seasons (as already mentioned above). We think that there is the possibility that spatial resolution of the FAPAR product may be too coarse. Negative observed GPP from eddy covariance data in November 2000 (Fig. 4) also indicate inadequate data processing in terms of the estimated daytime ecosystem respiration for calculating GPP by subtracting NEE from it or the gap-filling procedure for missing data.

5 Conclusions and implications

We present the first study that simultaneously assimilates latent heat fluxes as measured by the eddy covariance technique and satellite derived FAPAR, in CCDAS, here for the savanna site at Maun, Botswana. Simulated LHF and FAPAR show a reasonable seasonality for the case of assimilating the two data streams together. For assimilation of only one data stream, the seasonality of the other is slightly improved in

BGD
9, 3615–3643, 2012

comparison with observations. The optimization against both data streams leads to an average relative reduction in parameter uncertainty of more than 16.1 % for the 24 eco-hydrological parameters in CCDAS. This compares to only 9.2 % and 7.8 % for the single-data stream assimilation with LHF or FAPAR, respectively. Thus, assimilation of multiple observational data streams provides a greater potential to improve model accuracy and by reducing parameter uncertainty.

We further find that assimilation of either LHF oder FAPAR, or both, consistently constrain the key parameter in CCDAS that describes the water balance in semi-arid ecosystems, namely the maximum plant available soil moisture. Since FAPAR data from satellites is available with global coverage, this result offers (as demostrated by Kaminski et al., 2011) the potential of using FAPAR to constrain parameters related soil moisture, a quantity that is very difficult to observe on large spatial scales.

The approach of simultaneous assimilation of multi-data streams as presented here can be extended to include additional remote sensing products, for example using the surface soil moisture product from the Soil Moisture and Ocean Salinity (SMOS) mission (Kerr et al., 2001). This would allow a rigorous assessment of the consistency of multiple data streams (as done here for FAPAR and LHF). More importantly, the combined assimilation of FAPAR data with surface soil moisture from SMOS in CCDAS would lead to a more complete description of the hydrological properties than with just a dedicated soil moisture mission alone.

Supplementary material related to this article is available online at: http://www.biogeosciences-discuss.net/9/3615/2012/bgd-9-3615-2012-supplement.pdf.

Acknowledgements. This study was supported by the Postdoctoral Fellowships for Research Abroad, Japan Society for the Promotion of Science, the Ministry of Education, Science, Culture, Sports and Technology of Japan, and the QUEST programme funded by the Natural Environment Research Council, UK.

BGD
9, 3615–3643, 2012

The publication of this article is financed by CNRS-INSU.

References

Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water vapor exchange of forests: the EUROFLUX methodology, Adv. Ecol. Res., 30, 113–175, 2000.

Baldocchi, D. D., Falge, E., Gu, L., Olson, R., Hollinger, D. Y., Running, S. W., Anthoni, P., Bernhofer, Ch., Davis, K. J., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B. E., Lee, X., Malhi, Y., Meyers, T. P., Munger, J. W., Oechel, W. C., Paw U, K. T., Pilegaard, K. H., Schmid, P., Valentini, R., Verma, S., Vesala, T., Wilson, K. B., and Wofsy, S. C.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor and energy flux densities, B. Am. Meteorol. Soc., 82, 2415–2434, 2001.

Barbu, A. L., Calvet, J.-C., Mahfouf, J.-F., Albergel, C., and Lafont, S.: Assimilation of Soil Wetness Index and Leaf Area Index into the ISBA-A-gs land surface model: grassland case study, Biogeosciences, 8, 1971–1986, doi:10.5194/bg-8-1971-2011, 2011.

Braswell, B. H., Sacks, W. J., Linder, E., and Schimel D. S.: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, Glob. Change Biol., 11, 335–355, 2005.

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

Giering, R. and Kaminski, T.: Recipes for adjoint code construction, ACM Transactions on Mathematical Software, 24, 437–474, 1998.

BGD
9, 3615–3643, 2012

Gobron, N., Pinty, B., Aussedat, O., Chen, J. M., Cohen, W. B., Fensholt, R., Gond, V., Lavergne, T., Mélin, F., Privette, J. L., Sandholt, I., Taberner, M., Turner, D. P., Verstraete, M. M., and Widlowski, J.-L.: Evaluation of Fraction of Absorbed Photosynthetically Active Radiation Products for Different Canopy Radiation Transfer Regimes: Methodology and Results Using Joint Research Center Products Derived from SeaWiFS Against Ground-Based Estimations, J. Geophys. Res.-Atmos., 111, D13110, doi:10.1029/2005JD006511, 2006

Heimann, M.: The TM2 tracer model, model description and user manual, Technical Report No. 10, ISSN 0940-9327, Deutsches Klimarechenzentrum, Hamburg, 47 pp., 1995.

Ichii, K., Wang, W., Hashimoto, H., Yang, F., Votava, P., Michaelis, A. R., and Nemani, R. R.: Refinement of rooting depths using satellite-based evapotranspiration seasonality for ecosystem modeling in California, Agr. Forest Meteorol., 149, 1907–1918, 2009.

IPCC: Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK, 2007.

Jones, P., Osborn, T., Briffa, K., Folland, C., Horton, B., Alexander, L., Parker, D., and Rayner, N.: Adjusting for sampling density in grid-box land and ocean surface temperature time series, J. Geophys. Res., 106, 3371–3380, 2001.

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Consistent assimilation of MERIS FAPAR and atmospheric CO_2 into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences Discuss., 8, 10761–10795, doi:10.5194/bgd-8-10761-2011, 2011.

Kleidon, A. and Heimann, M.: Simulating root carbon storage with a coupled carbon-water cycle root model, Phys. Chem. Earth, 21, 499–502, 1998.

Kerr, Y. H., Waldteufel, P. Wigneron, J.-P., Martinuzzi J., Font, J., and Berger M.: Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. Geoscience and Remote Sensing, IEEE T. Geosci. Remote, 39, 1729–1735, 2001.

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

Knorr, W. and Heimann, M.: Uncertainties in global terrestrial biosphere modeling, Part I: a comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, Global Biogeochem. Cy., 15, 207–225, 2001.

BGD
9, 3615–3643, 2012

Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Carbon Cycle Data Assimilation with a Generic Phenology Model, J. Geophys. Res. Atmos, 115, G04017, doi:10.1029/2009JG001119, 2010.
Knorr, W. and Kattge, J.: Inversion of terrestrial biosphere model parameter values against eddy covariance measurements using Monte Carlo sampling, Global Change Biol., 11, 1333–1351, 2005.
Lloyd, J., Kolle, O., Veenendaal, E., Arneth, A., and Wolski, P.: SAFARI 2000 Meteorological and Flux Tower Measurements in Maun, Botswana, 2000. Data set, available at: http://daac.ornl.gov/ from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA doi:10.3334/ORNLDAAC/760, 2004.
Mantlana, B. K.: Physiological characteristics of two forms of *Colophospermum mopane* growing on Kalahari sand. M.S. Thesis, University of Natal, Durban, South Africa, 2002.
Moncrieff, J. B., Massheder, J. M., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S, Sogaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide, J. Hydrology, 188–189, 589–611, 1997.
Nijssen, B., Schnur, R., and Lettenmaier, D.: Retrospective estimation of soil moisture using the VIC land surface model, 1980–1993, J. Climate, 14, 1790–1808, 2001.
Rayner, P., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial Carbon fluxes from a Carbon Cycle Data Assimilation System (CCDAS), Global Biogeochem. Cy., 19, GB2026, doi:10.1029/2004GB002254, 2005.
Schenk, H. J. and Jackson, R. B.: The global biogeography of roots, Ecological Monographs, 72, 311–328, 2002.
Scholze, M.: Model studies on the response of the terrestrial carbon cycle on climate change and variability, Examensarbeit, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 2003.
Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic CCDAS simulations, J. Geophys. Res., 112, D17305, doi:10.1029/2007JD008642, 2007.
Veenendaal, E. M., Kolle, O., and Lloyd, J.: Seasonal variation in energy fluxes and carbon dioxide exchange for a broad-leaved semi-arid savanna (Mopane woodland) in Southern Africa, Global Change Biol., 10, 309–317, 2004.

BGD
9, 3615–3643, 2012

Veenendaal, E., Mantlana, K. B., Pammenter, N. W., Weber, P., Huntsman-Mapila, P., and Lloyd, J.: Growth form and seasonal variation in leaf gas exchange of *Colophospermum mopane* savanna trees in northwest Botswana, Tree Physiol., 28, 417–424, 2008.

Wang, G.: Agricultural drought in a future climate: Results from 15 global climate models participating in the IPCC 4th Assessment, Clim. Dynam., 25, 739–753, 2005.

Williams, M,, Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Glob. Change Biol., 11, 89–105, 2005.

BGD

9, 3615–3643, 2012

Table 1. Cost function contributions from parameters (Param.), latent heat flux (LHF), and fraction of absorbed photosynthetically active radiation (FAPAR) as well as the total cost function value and final value of the gradient for the total cost function.

Run	Cost function				
	Param.	LHF	FAPAR	Total	Gradient
Prior	0	470	1825	470 (LHF), 1825 (FAPAR), 2295 (Combined)	n.d.*
Experiment 1 (LHF)	11	302	1576**	313	1.5×10^{-2}
Experiment 2 (FAPAR)	19	1199**	13	32	5.4×10^{-7}
Experiment 3 (Combined)	20	746	141	908	4.6×10^{-6}

* n.d. is no data, **not counted for total cost function.

BGD
9, 3615–3643, 2012

Table 2. List of parameters in prior run, and posterior runs assimilating LHF, FAPAR, and combination of LHF and FAPAR. Uncertainty reduction (unc. red.) is calculated as posterior minus prior uncertainty divided by prior uncertainty. Top rows: physiology, middle: phenology, bottom: energy and water budgets. Units of parameters are: V_{max} in $\mu mol(CO_2)\,m^{t2}\,s^{-1}$, k in $\mu mol(air)\,m^{-2}\,s^{-1}$, $\alpha_{\Gamma,T}$ in $\mu mol(CO_2)mol(air)^{-1}{}^{\circ}C^{-1}$, K_C in $\mu mol(CO_2)mol(air)^{-1}$, K_O in $mol(O_2)mol(air)^{-1}$, activation energies E in J mol^{-1}, τ_W in days, C_{W0} in mm hour^{-1}, W_{max} in mm, others unitless. Prior uncertainty represents one standard deviation, except for the log-normally distributed parameters denoted by (*), for which the analogous difference between mean and upper 67 %-tile is given.

Num.	PFT	Parameter	Prior		Experiment 1 (LHF)		Experiment 2 (FAPAR)		Experiment 3 (Combined)	
			value	unc.	value	unc. red. [%]	value	unc. red. [%]	value	unc. red. [%]
1	2	V_{max}^{25}	90	18	75	2	78	3	34	42
2	10	V_{max}^{25}	8	1.6	6	19	8	2	6	56
3	2	$a_{J,V}$	1.99	0.10	1.99	0	1.99	0	2.01	0
4	10	k^{25}	140	28	140	0	140	0	140	0
5	All	E_{Rd}	45 000	2250	45 021	0	44 992	0	45 088	0
6	All	E_{Vmax}	58 520	2926	58 429	1	58 430	0	54 748	0
7	2	E_{KO}	35 948	1797	35 924	0	35 952	0	35 732	0
8	2	E_{KC}	59 356	2967	59 228	1	59 367	0	60 958	0
9	10	E_k	50 967	2548	50 966	0	50 966	0	50 959	0
10	2	α_q	0.280	0.04	0.280	0	0.282	0	0.288	4
11	10	α_i	0.040	0.002	0.040	0	0.040	0	0.041	0
12	2	K_C^{25}	460	23	463	0	463	0	470	1
13	2	K_O^{25}	0.33	0.02	0.33	0	0.33	0	0.33	0
14	2	$\alpha_{\Gamma,T}$	1.70	0.09	1.70	0	1.69	0	1.68	1
15	All	Λ_{max}	5.00	0.25	5.20	13	5.03	1	5.05	3
16	All	ξ	0.50	0.10	0.50	0	0.50	3	0.61	14
17	2	τ_W^*	30	15	10	43	186	43	94	87
18	10	τ_W^*	30	15	11	36	22	37	14	56
19	2	fci_{C3}	0.650	0.065	0.589	10	0.653	22	0.695	22
20	10	fci_{C4}	0.370	0.037	0.360	4	0.371	0	0.379	4
21	All	C_W	0.500	0.005	0.500	0	0.500	0	0.502	0
22	All	h_0	0.490	0.005	0.490	0	0.490	0	0.488	0
23	All	$\hat{h}$	0.960	0.010	0.959	0	0.961	0	0.953	0
24	All	W_{max}	1500	1500.0	332	90	86	75	129	95

BGD
9, 3615–3643, 2012

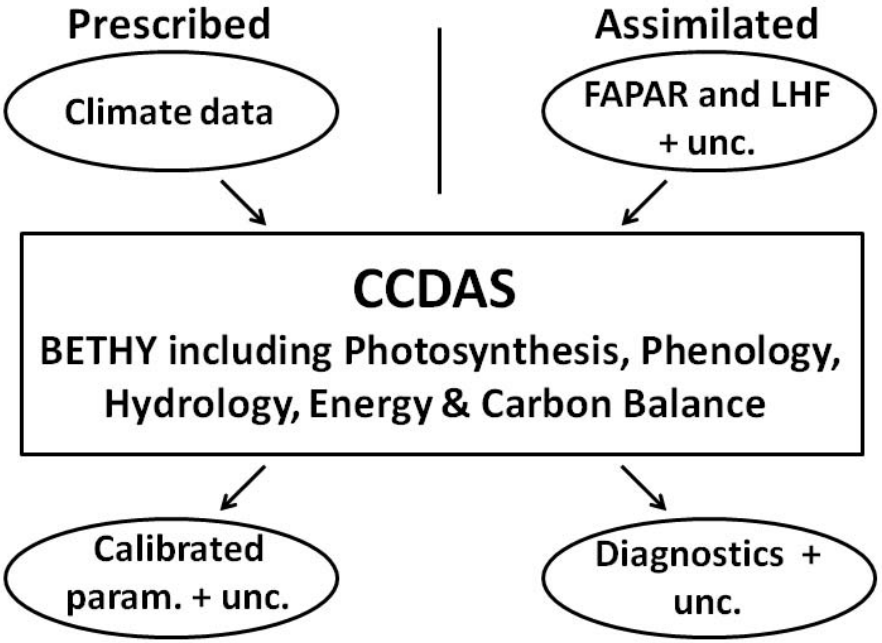

Fig. 1. Schematic diagram of the CCDAS structure. Ovals represent input and output data, and boxes represent calculation steps. Diagnostics are quantities of interest such as carbon fluxes computed by CCDAS. Unc. stands for uncertainty and param. for parameters.

BGD
9, 3615–3643, 2012

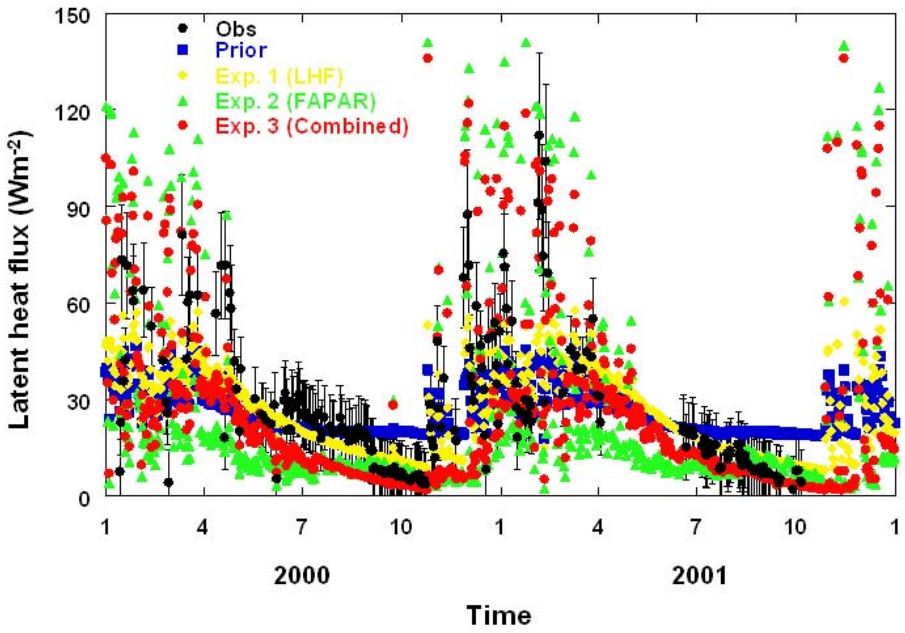

Fig. 2. Observed and simulated latent heat flux (LHF; W m^{-2}) for the years 2000–2001. The error bar for observed LHF (Obs) represents the data uncertainty used in CCDAS. Prior, Posterior LHF, Posterior FAPAR, and Posterior combined are prior, posterior runs with LHF, FAPAR, and combined of LHF and FAPAR, respectively. Root mean square errors (RMSEs) of simulated LHF against observation are 17.2, 14.6, 26.5 and 21.5 W m^{-2} for prior, experiment 1, 2 and 3, respectively.

BGD
9, 3615–3643, 2012

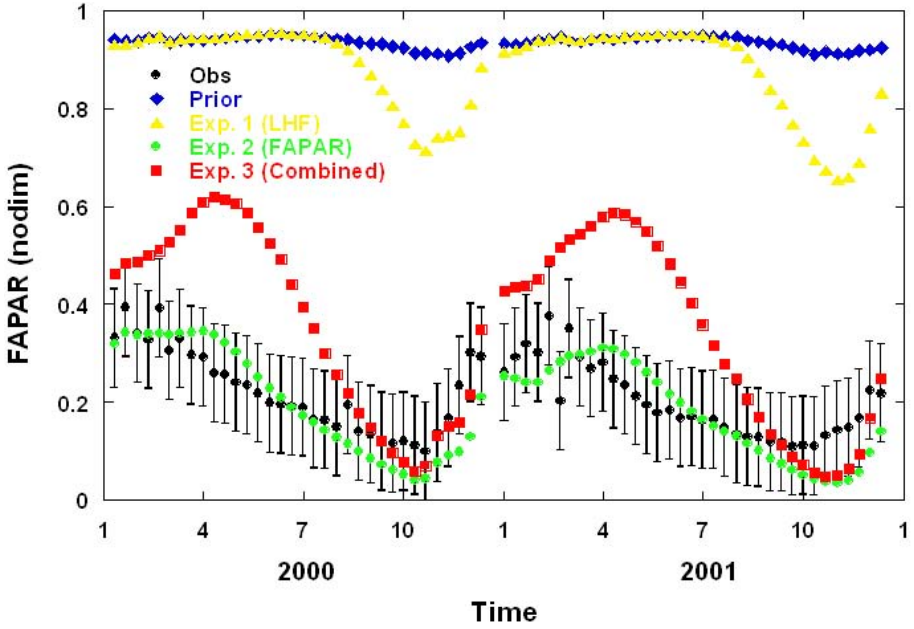

Fig. 3. Observed and simulated fraction of absorbed photosynthesically active radiation (FAPAR; nodim) for the years 2000–2001. Observed FAPAR is derived from the SeaWiFS instrument from the National Aeronautics and Space Administration (NASA). The error bar of observed FAPAR (Obs) represents the data uncertainty used in CCDAS. Prior, Posterior LHF, Posterior FAPAR, and Posterior combined are prior, posterior runs with LHF, FAPAR, and combined of LHF and FAPAR, respectively. Root mean square errors (RMSEs) of simulated FAPAR against observation are 0.727, 0.676, 0.062 and 0.202 for prior, experiment 1, 2 and 3, respectively.

BGD
9, 3615–3643, 2012

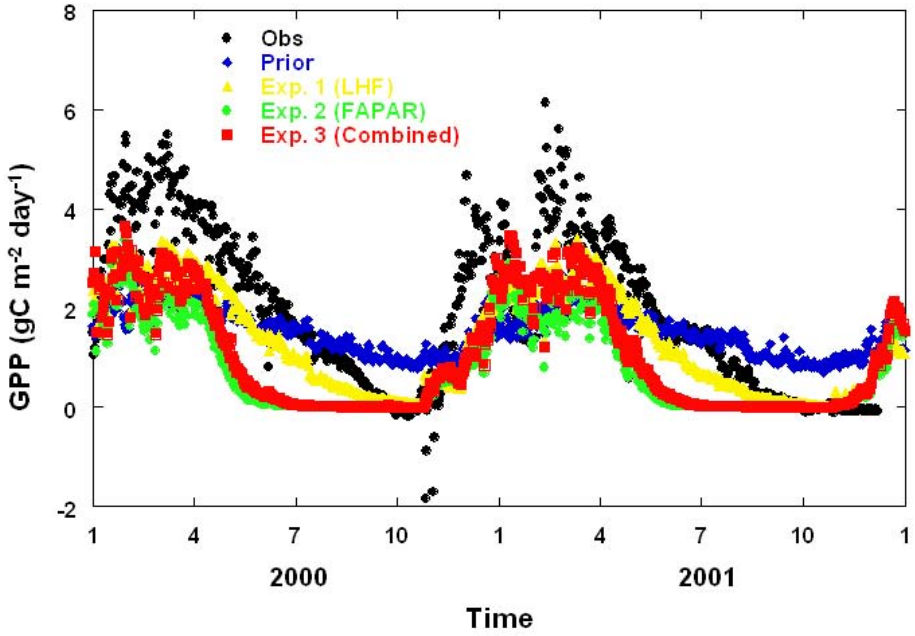

Fig. 4. Observed and simulated gross primary production (GPP) for the years 2000–2001. Observed GPP (Obs), based on eddy covariance data, is estimated by subtracting net ecosystem exchange (NEE) from ecosystem respiration during daytime. Daytime ecosystem respiration is estimated from the relationship between nighttime ecosystem respiration and soil temperature. Prior, Post LHF, Post FAPAR, and Post combined are prior, posterior runs with LHF, FAPAR, and combined of LHF and FAPAR, respectively. Root mean square errors (RMSEs) of simulated GPP against observation-based values are 1.40, 0.95, 1.61 and 1.14 g C m^{-2} day^{-1} for prior, experiment 1, 2, and 3, respectively.

BGD
9, 3615–3643, 2012

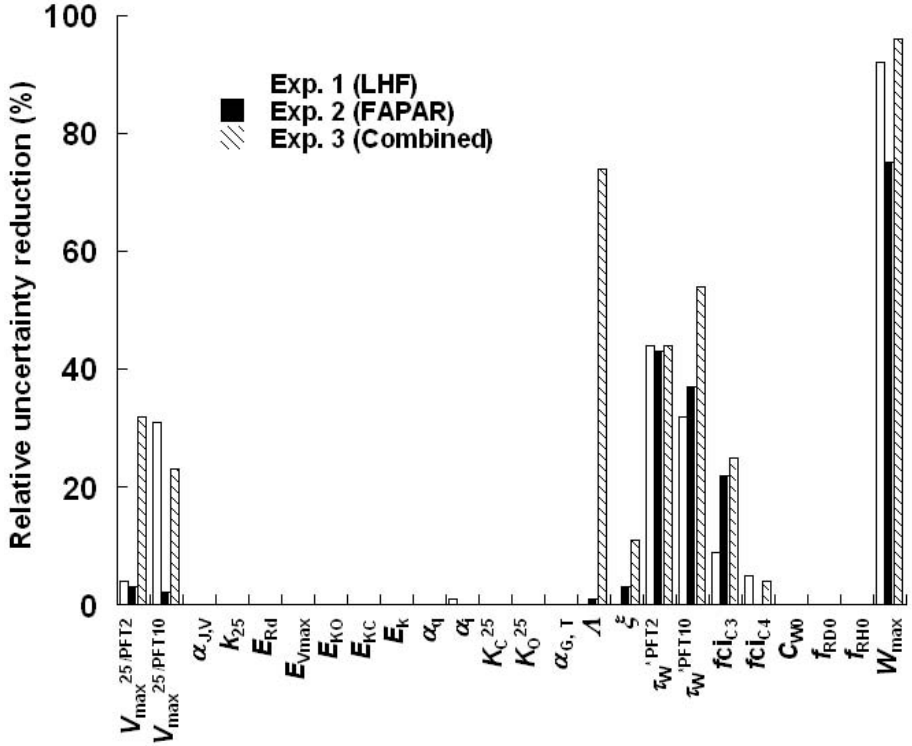

Fig. 5. Relative reduction (%) in parameter uncertainty after optimization.

BGD
9, 3615–3643, 2012