



The potential benefit of using forest biomass data in addition to carbon and water flux measurements to constrain ecosystem model parameters: Case studies at two temperate forest sites

T. Thum^{a,*}, N. MacBean^b, P. Peylin^b, C. Bacour^c, D. Santaren^b, B. Longdoz^d, D. Loustau^e, P. Ciais^b

^a Finnish Meteorological Institute, Climate Research, P.O. Box 503, FI-00101 Helsinki, Finland

^b Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

^c Noveltis, 153 Rue du Lac, F-31670 Labège, France

^d Gembloux Agro-Bio Tech, University of Liège, Belgium

^e INRA, UR1263 EPHYSE, Villenave d'Ornon, France

ARTICLE INFO

Article history:

Received 12 May 2015

Received in revised form

29 November 2016

Accepted 4 December 2016

Keywords:

Model data fusion

ORCHIDEE

Eddy covariance

ABSTRACT

Biomass as a resource, and as a vulnerable carbon pool, is a key variable to diagnose the impacts of global changes on the terrestrial biosphere, and therefore its proper description in models is crucial. Model-Data Fusion (MDF) or data assimilation methods are useful tools in improving ecosystem models that describe interactions between vegetation and atmosphere. We use a MDF method based on a Bayesian approach, in which data are combined with a process model in order to provide optimized estimates of model parameters and to better quantify model uncertainties, whilst taking into account prior information on the parameters. With this method we are able to use multiple data streams, which allows us to simultaneously constrain modeled variables at site level across different temporal scales. In this study both high frequency eddy covariance flux measurements of net CO₂ and evapotranspiration (ET), and low frequency biometric measurements of total aboveground biomass and the annual increment (which includes all compartments), are assimilated with the ORCHIDEE model version "AR5" at a beech (Hesse) and a maritime pine (Le Bray) forest site using four to five years of flux data and nine years of biomass data. When assimilating the observed aboveground annual biomass increment (AGB.inc) together with net CO₂ and ET flux, the RMSE of modelled AGB.inc was reduced from the a priori estimates by 37% at Hesse and 69% at Le Bray, without reducing the fit to the net CO₂ and ET that can be achieved when assimilating flux data alone. Assimilating biomass increment data also provides insight in the performance of the allocation scheme of the model. Comparison with detailed site-based measurements at Hesse showed that the optimization reduced positive biases in the model, for example in fine root and leaf production. We also investigated how to use stand-scale total aboveground biomass in optimization (AGB.tot). However, this study demonstrated that assimilating AGB.tot measurements in the ORCHIDEE-AR5 model lead to some inconsistencies, particularly for the annual dynamics of the AGB.inc, partly because this version of the model lacked a realistic representation of forest stand processes including management and disturbances.

© 2016 Elsevier B.V. All rights reserved.

* Corresponding author at: Finnish Meteorological Institute, Climate Research, P.O. Box 503, FI-00101 Helsinki, Finland.

E-mail addresses: tea.thum@fmi.fi (T. Thum), natasha.macbean@lsce.ipsl.fr (N. MacBean), philippe.peylin@lsce.ipsl.fr (P. Peylin), cedric.bacour@lsce.ipsl.fr (C. Bacour), diego.santaren@lsce.ipsl.fr (D. Santaren), Bernard.Longdoz@ulg.ac.be (B. Longdoz), Denis.Loustau@bordeaux.inra.fr (D. Loustau), philippe.ciais@lsce.ipsl.fr (P. Ciais).

1. Introduction

It is of critical importance to understand how the terrestrial carbon cycle will respond to changing climatic conditions and anthropogenic activity. The global terrestrial carbon cycle has been intensively studied, but large uncertainties in the model predictions of major sources and sinks remain (Ciais et al., 2013). The inter-model spread using different coupled climate-carbon cycle models demonstrates that there are still large uncertainties in carbon flux and stock projections by the end of this century (Friedlingstein et al.,

2014), with some models predicting that the terrestrial biosphere will become a source of carbon to the atmosphere, while others predict an increase in the carbon storage (Ahlinström et al., 2012). Whilst one source of uncertainty is the result of incorrect or incomplete biophysics, in particular regarding the models' response to elevated CO₂ (Zaehle et al., 2014), another arises from poorly calibrated model parameters (Schwalm et al., 2010). A third issue is related to difficulty and uncertainties in accounting for the history of land cover dynamics and disturbances (natural and human induced) in the model initialization, which directly controls the vegetation and soil states at the start of a simulation (Hurt et al., 2010).

Model-Data Fusion (MDF) is a statistically rigorous approach that can be used to optimize model parameters, given different observations and prior information on the parameters. MDF has been applied extensively to optimize ecosystem models, including those which form the land surface component of global Earth System Models, with in situ net CO₂ and evapotranspiration (ET) flux data (Braswell et al., 2005; Wang et al., 2007; Williams et al., 2009; Carvalhais et al., 2010; Kuppel et al., 2012; Santaren et al., 2014), as well as satellite NDVI/fAPAR data (Knorr et al., 2010; Bacour et al., 2015; MacBean et al., 2015) and atmospheric CO₂ concentration data (Kaminski et al., 2012; Peylin et al., 2016; Schürmann et al., 2016). Typically flux tower data bring information on the “fast” processes that control fluxes of photosynthesis, respiration and evapotranspiration on a diurnal to seasonal timescale, whereas their potential for constraining models at inter-annual timescales is lower (Wang et al., 2012; Santaren et al., 2014). Additional measurements of other ecosystem variables (e.g. above and below-ground carbon stocks) are often available at flux tower locations (Curtis et al., 2002) and can be used for optimizing the “slower” carbon processes in ecosystem models, such as carbon allocation, turnover and mortality. However, long term carbon stock and stock increment data have not been used as often in model optimization studies to date, partly because they involve longer time scales where initial conditions (Carvalhais et al., 2008, 2010) and disturbance history adds uncertainties (Thornton et al., 2002).

The use of multiple data streams in MDF has not yet been extensive, but some studies have investigated the added benefit different data sources, including some that use ground-based data that are related to canopy-scale biomass. Keenan et al. (2012) used 18 years of NEE flux data from the Harvard forest combined with ecological measurements, such as phenological observations of leaf budburst, litter data, carbon pools in different biomass compartments to constrain the carbon cycle-related parameters of the Föbäa model. Similarly Wu et al. (2011) used micrometeorological flux data together with biomass data and LAI observations. Bacour et al. (2015) examined the complementarity of in situ flux and fAPAR (fraction of the Absorbed Photosynthetically Active Radiation) data for optimizing parameters related to photosynthesis and phenology in the ORCHIDEE model. Most of these studies reported that using additional information provided an extra constraint that enabled a greater number of parameters to be independently resolved by the optimization. The importance of different data streams can be seen in their relative influence on the reduction of uncertainties on the optimized parameters (Wu et al., 2011), but this is also dependent on the uncertainties associated to each data stream, possible biases in the data and how many observations they contain as well as possible model-data biases associated to each data stream (see MacBean et al., 2016 for a review of the challenges involved with multiple data stream assimilation).

MDF can also be used to study inconsistencies between data-streams, between a given model-structure and real-world observations, and also to identify processes that are not well understood in models and give perspective to further measurement needs (e.g. Keenan et al., 2013). Using several different data

streams helps in constraining a larger parameter space (Kaminski et al., 2012) and enables a better evaluation of the model, possibly even highlighting areas where model needs improving and thus helping to set priorities for future model developments (Rayner, 2010). If the model cannot be optimized to match the observations within prescribed uncertainties of an unbiased data stream it shows that the model structure may be insufficient to describe the processes that represent the variability of observed quantities, and thus requires further modification.

In recent years large-scale maps of biomass have become available (Saatchi et al., 2011; Baccini et al., 2012; Thurner et al., 2014; Santoro et al., 2015; Avitabile et al., 2016) and possess a high potential to deepen our knowledge of ecosystem functioning (e.g. Carvalhais et al., 2014). In addition, the planned ESA BIOMASS satellite mission (http://www.esa.int/Our_Activities/Observing_the_Earth/The_Living_Planet_Programme/Earth_Explorers/Future_missions/Biomass) has the potential to provide us with more accurate global coverage of forest biomass and forest height estimates. These data will be very useful in constraining the slower processes in land surface models (LSMs) at regional scales. However, integrating in situ flux and biomass data into an ecosystem model to constrain both “fast” and “slow” carbon processes imposes a challenge for a MDF approach. Most global scale models use a steady state assumption for forest growth that results in an overestimation of the simulated tree biomass for young forests (Pietsch and Hasenauer, 2006; Ciais et al., 2008), as the steady state assumption corresponds to old-growth forests that hold maximum carbon stocks. Younger forests are usually carbon sinks because the trees are growing and store carbon as they age.

In order to use biomass data in an optimization, realistic simulation of forest growth, from planting (or last disturbance) to its current age, is required, which might impose a strong constraint in terms of computing time, especially when using Monte Carlo algorithm in an optimization. Usually, land surface models use rather complex equations (with many parameters) to describe photosynthesis and organic matter decomposition, but very simple equations (with few parameters) to describe biomass dynamics (Purves and Pacala, 2008). In these models, processes such as stand level recruitment, competition and specific mortality processes including disturbances are not represented, and wood biomass is considered to be a well-mixed pool (Bolin and Rodhe, 1973) whose mass balance results from the input of a fraction of net primary production (NPP) that is allocated to wood, and removal assumed to be a constant fraction of the pool mass (Friedlingstein et al., 1999).

In this study we investigated how to use aboveground biomass data, both the total stock and the annual increment for all biomass compartments, in MDF with a global-scale process-based model ORCHIDEE (ORGanizing Carbon and Hydrology In Dynamic Ecosystems) (Krinner et al., 2005), version AR5. Note that ORCHIDEE-AR5 is also the surface component of an Earth System Model (IPSL – Institut Pierre Simon Laplace) used to make future climate predictions, and the AR5 version was used for the last IPCC Assessment Report (Ciais et al., 2013). The challenges surrounding the optimization of a global LSM using different time scales of observation data, as discussed above, required specific optimization strategies to be tested first at site-level. The main objective therefore is to investigate different approaches to combine micrometeorological in situ flux and biomass data within a MDF framework. The study is conducted for two different forest sites separately (beech and pine forests). In this context, our study aims to investigate the following questions:

- Q1: What are the challenges linked to using flux data, annual total aboveground biomass and aboveground biomass increment data together in the same optimization at a given site?

- Q2: How does the assimilation of yearly biomass increment data help to constrain carbon allocation parameters in a process-based model compared to just using flux data?
- Q3: What do we gain when we add optimization of residence time (turnover rate) with total aboveground biomass data after optimization of fast carbon processes with fluxes and aboveground biomass increment?

First we study the benefits of including aboveground biomass increment in the optimization with the flux data and then study the feasibility of optimizing with total biomass in a second step (i.e. where each data stream is assimilated separately). In this approach the first step is dedicated to optimizing the fast processes that correspond to a large number of relevant parameters, and the second step to optimizing the slow processes associated with a smaller number of parameters while keeping the parameter values for the fast processes that were inferred in the first step.

2. Materials and methods

Our study sites include the Hesse beech forest and the Le Bray maritime pine forest, both of which are located in France. The optimization was performed at both sites separately. We also compared the optimized model results at Hesse to another dataset at the site (Granier et al., 2008) and performed a future scenario run to assess the impact that optimized parameter values will have on the future deterministic trajectory of carbon storage at the two forests (and their associated uncertainties).

2.1. Site descriptions and measurements

2.1.1. Broadleaf deciduous “Hesse”

Hesse is a beech (*Fagus sylvatica*) forest located in the temperate region of northeastern France (48°40'N, 7°05'E, 300 m a.s.l.). The mean annual rainfall at the site is 820 mm and annual mean temperature is 9.2 °C (Granier et al., 2000). The forest was 40 years old (± 5 years) on average in 2005 (Granier et al., 2008) and thus was planted around 1965. The Hesse forest is a young growth forest, with thinnings taking place every 5–6 years. Thinnings were done at the end of 1995, in March 2002 and at the end of 2005; about 25% of the basal area was removed each time (Granier et al., 2008). The value of leaf area index (LAI) has changed between 4.6 and 7.6 m² m⁻², with a sharp decline occurring after each thinning had taken place followed by a rapid recovery in just a few years (Granier et al., 2008). The soil type is luvisol/stagnic luvisol (Granier et al., 2000).

Site level data (fine root and leaf production, autotrophic respiration and net primary production, NPP) from study by Granier et al. (2008) was used for comparison with the optimized model results for years 1995–2005. The observed fine root production was estimated from allometric relationships and the leaf production from litter trap measurements. Autotrophic respiration was measured using the chamber method. NPP is defined as GPP minus the autotrophic respiration; thus it is the net amount of carbon assimilated during photosynthesis that is available to the plants after growth and maintenance respiration. We used the NPP estimated based on micrometeorological flux measurements with the estimate for autotrophic respiration from chamber measurements.

2.1.2. Needleleaf evergreen “Le Bray”

Le Bray is a maritime pine (*Pinus pinaster*) forest located in Southern France (44°43'N, 0°46' W, 300 m a.s.l.) that was planted in 1970. The annual rainfall is 972 mm (Jarosz et al., 2008) and annual mean temperature is 12.9 °C (Medlyn et al., 2002). A storm in 1999 killed some trees and in 2005 some thinning was performed at the site. The LAI ranges from 2.6 m² m⁻² in late winter to 3.1 m² m⁻² in

early autumn (Berbigier et al., 2001). The climate has strong intra-seasonal variability, with high precipitation during winter and dry periods during summer when soil drought often occurs (Ogée et al., 2003), but the forest is able to photosynthesize year round due to relatively mild climate and lack of limiting factors. The soil type is a sandy and hydromorphic podsol (Berbigier et al., 2001).

2.1.3. Measurements at the sites

Ecosystem scale net CO₂, water and energy flux measurements using the eddy covariance technique started at both sites in 1996. The data consist of net CO₂ and ET fluxes and meteorological data that are processed as described in Aubinet et al. (2000), Reichstein et al. (2005) and Papale et al. (2006). The measured net ecosystem CO₂ flux, Net Ecosystem Exchange (NEE), is a combination of two fluxes with opposite signs, the total ecosystem respiration (TER) and gross primary production (GPP). The flux-partitioning method of Reichstein et al. (2005) separates NEE using night-time observations and a short-term temperature response function for the respiration, and therefore provides GPP and TER from NEE measurements. In this study we use the GPP and TER, and not the NEE, data together with ET (and biomass – see below) in the optimization.

At Hesse the stand-scale biomass was estimated from circumference measurements using allometric relationships established at the site (Grote et al., 2011). Biomass was estimated using allometric equations and circumference measurements each year between 1995 and 2005, once a year in regular years and twice a year when thinnings took place (Granier et al., 2008). Similar biomass measurements were made at Le Bray for each year from 1996 to 2007, following the allometric relationships developed for the site (Porté et al., 2002). The annual aboveground biomass increment (AGB.inc) estimate is based only on the measured natural circumference growth and natural mortality of the forest and therefore follows a positive trend, whereas the total aboveground biomass (AGB.tot) takes the biomass removed from the site into account, and thus under this definition, the values fluctuate from year to year (i.e., one year can be lower than the previous year). There is no significant natural mortality in these two forests.

2.2. Model description

2.2.1. The ORCHIDEE model

We use the ecosystem model ORCHIDEE, version AR5. It is a process-based model describing the exchange of carbon, water and energy between the atmosphere and vegetation as well as carbon and water pool dynamics (Krinner et al., 2005). The model has primarily been developed to run at global scale coupled with a climate model, but in this study only site level runs were performed, with fixed vegetation and climatic forcing measured at the site.

In ORCHIDEE the vegetation is discretized into Plant Functional Types (PFTs) so that the beech forests have parameter values common to the *deciduous temperate forest* PFT and the maritime pine forest follows the same equations but with different parameter values that apply for the *temperate coniferous forest* PFT for most of the processes. The two PFTs differ in their parameter values and the equations that define their phenology.

The fast processes in ORCHIDEE are calculated at a half-hourly time step including the assimilation and the respiration of carbon, the hydrology and the energy cycles. The photosynthesis is based on the mechanical description of Farquhar et al. (1980) with the stomatal conductance following the Ball-Woodrow-Berry model (Ball et al., 1987). The autotrophic respiration consists of growth and maintenance respiration. The maintenance respiration is a function of temperature and biomass, whereas the growth respiration is a prescribed fraction of the assimilated carbon. The assimilated carbon is allocated into different biomass compartments (see below) at a daily time step. The heterotrophic respiration originates from

litter pools located above- and belowground and three soil carbon pools, as is described in Parton et al. (1988). Six litter pools exist in ORCHIDEE: structural, metabolic and woody litter; all have above and below ground compartments (Zaehle and Friend, 2010). The model equations are described in several papers and for the carbon cycle we recommend Krinner et al. (2005), with an update in Santaren et al. (2014). Appendix A provides the equations of the fast processes that are optimized in this work, while the next section and Appendix B gives a specific description of the carbon allocation scheme that is crucial for this study in more detail.

2.2.2. Biomass pools and carbon allocation

A schematic showing the different pools and the parameters that determine the division of NPP in the ORCHIDEE model is shown in Fig. 1. The biomass of vegetation is divided into eight different compartments in ORCHIDEE and these include above- and belowground sapwood and heartwood (four compartments altogether), leaves, fine roots, fruits and carbohydrate reserve. The fruit compartment includes plant parts with reproductive functions, such as flowers and fruits. The carbohydrate reserve enables the production of leaves at leaf onset; this compartment is only present for deciduous PFTs. A part of the sapwood is turned into heartwood by a constant turnover rate. The separation of biomass between heartwood and sapwood does not influence the autotrophic respiration rates.

The allocation of assimilated carbon in ORCHIDEE is based on Friedlingstein et al. (1999), and assumes optimal use of resources depending on the environmental conditions. Carbon is allocated to sapwood, fine roots or leaves, after maintenance respiration is subtracted from GPP. If the plant is light limited, it allocates carbon to sapwood. If there is soil moisture limitation, more carbon is allocated to fine roots. The equations governing the allocation are shown in Appendix B. There are two parameters controlling the allocation to fine roots and sapwood, r_0 and s_0 , respectively. These two parameters were optimized in this study (see Section 2.5.2). Sapwood is further split into above and belowground compartments, according to the age of the forest. This separation is determined by the equation

$$F_{\text{above}} = S_{\text{init}} + 0.6 * (1 - \exp(-A_{\text{age}}/d_{\text{alloc}})) \quad (1)$$

where F_{above} is the part of the NPP allocated to sapwood that will be located to AGB compartments, S_{init} the fraction of sapwood allocation aboveground at $A_{\text{age}} = 0$, A_{age} is the age of the forest (in years) and d_{alloc} is a time constant that defines the inertia of the age-dependent allocation of sapwood to AGB. The parameters S_{init} and d_{alloc} were optimized. In addition to these four allocation parameters, the joint turnover rate of all biomass compartments (i.e. the residence time T_{res}) was also optimized. T_{res} is used to define biomass loss to litter in ORCHIDEE. The inverse of T_{res} defines a constant mortality rate given that the modelled biomass removal in ORCHIDEE assumes a first order kinetics (Krinner et al., 2005). No disturbances such as fire, drought, insect attacks, thinning that would lead to an abrupt decrease of forest biomass (or increase in mortality) are described in the AR5 version of the model. The five parameters related to the dynamics of biomass are shown with their prior values in Table 1.

2.2.2.1. New LAI scheme. In a more recent forest module of ORCHIDEE, the development of LAI was slowed down for young forests (Bellassen et al., 2010). Inspired by this method, we tested a delaying of the development of LAI in young forests, which we hereafter refer to as the “new LAI scheme”. For temperate forests the delay was imposed over the first 15 years. In these years the maximum possible LAI ($LAI_{\text{max_act}}$) was the default maximum LAI

(LAI_{max}) decreased by the square root of the age (t_{age}) divided by the time used for the delay ($t_{\text{lai_limit}}$):

$$LAI_{\text{max_act}} = \sqrt{\frac{t_{\text{age}}}{t_{\text{lai_limit}}}} \cdot LAI_{\text{max}} \quad (2)$$

We tested this scheme as part of this study given its relevance for simulating aboveground biomass.

2.3. Model data fusion (MDF) method

We applied a statistical optimization, referred to as MDF, of the model parameters using the data described in Section 2.1.3 in which all sources of uncertainties are taken into account: measurement errors, model structural errors and parameter errors.

The MDF method uses a Bayesian formulation, which allows prior information about the selected parameters (see Section 2.5.1) to be updated based on new information (observations) (Tarantola, 2005). We assumed Gaussian errors with truncated PDFs for the parameter errors in order to account for physical bounds (see Section 2.5.1). The optimization follows the minimization of a cost function that measures the misfit between the observations and the corresponding model variables as well as between the estimated parameters and a priori values:

$$J(\mathbf{x}) = \sum_i \frac{1}{2} \left[\frac{1}{N_i} [(\mathbf{Y}_i - \mathbf{M}_i(\mathbf{x}))^t \mathbf{R}^{-1} (\mathbf{Y}_i - \mathbf{M}_i(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_b)^t \mathbf{P}_b^{-1} (\mathbf{x} - \mathbf{x}_b)] \right], \quad (3)$$

where i stands for different data streams, \mathbf{x} is the parameter vector, \mathbf{Y} is the observation, N_i is the number of observations of each data stream, $\mathbf{M}_i(\mathbf{x})$ is the model output and \mathbf{x}_b is the vector with a priori parameter values. \mathbf{R} and \mathbf{P}_b are the error covariance matrices for the prior uncertainties on observations and parameters, respectively. As in most studies, both of these matrices are assumed to be diagonal, given the difficulties in properly defining cross-parameter or cross-observation error covariances. For the micrometeorological flux data, daily mean values were used in the optimization, thus circumventing a complicated treatment of the uncertainties for the half hourly data, which studies have shown have a Laplacian distribution (Hollinger and Richardson, 2005; Lasslop et al., 2008). This approach follows Richardson et al. (2010), who assumed that random errors in half-hourly eddy covariance data would be approximately normal when integrated over a day. Daily flux observations were used, with a further smoothing (15-day running average) applied. This allows us to focus the optimization on time scales ranging from approximately bi-weekly to seasonal and inter-annual variability.

Following the study of Santaren et al. (2014) (who also used the ORCHIDEE model), we used an implementation of the genetic algorithm (GA) (Goldberg, 1989), which is a “global search” method, to perform the inversions. This Monte Carlo type of approach proved to be more efficient than a gradient based method to find the global minimum of J and less prone to getting trapped in local minima (see Santaren et al., 2014 for more details). The GA uses the principles of the genetics and natural selection to perform a stochastic search over the whole parameter space (Goldberg, 1989; Haupt and Haupt, 2004), and following Gaussian truncated PDFs. In our set-up a population of 30 chromosomes was used with 40–100 iterations, and at each iteration 80% of the new chromosomes were created by “mating”, the rest by “mutations”. This set-up was determined in the Santaren et al. (2014) study which tested multiple configurations and decided on the best based on the convergence of the system (i.e. the point after which cost function does not change and the minimum has been reached). The convergence of the cost function was visually checked and confirmed within the 40–100 iterations.

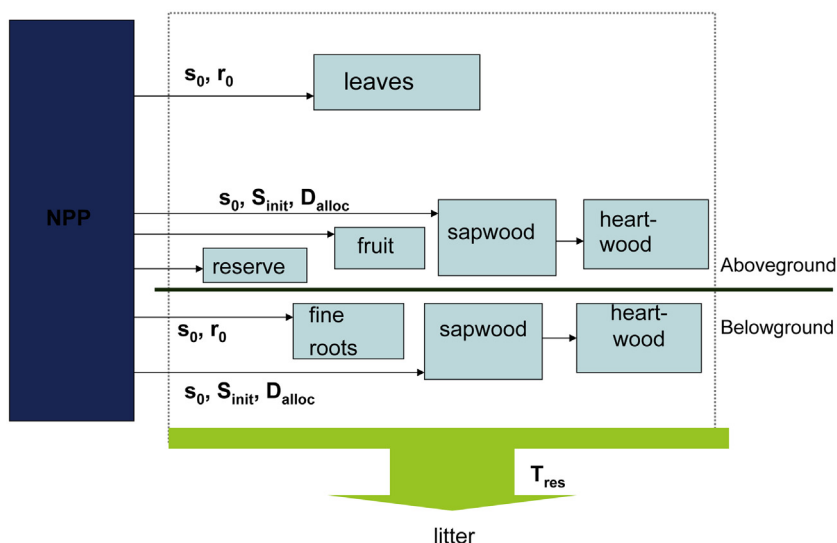


Fig. 1. Distribution of NPP into the different biomass pools of ORCHIDEE. Next to the arrow it is written which parameters influence the amount of NPP transported to the respective pool. The gray box in dotted line surrounds the biomass pools that are affected by the residence time.

Table 1
Optimized parameters. The prior value is given first for Hesse and then to Le Bray, if the value is different.

Parameter	Description	Units	Prior value	Parameter set
Photosynthesis				
$V_{c(max)}$	Maximum carboxylation rate	$\mu\text{mol m}^{-2} \text{s}^{-1}$	55.0/35.0	P1
$G_{S,slope}$	Slope in the stomatal conductance equation (Ball-Berry)	–	9.0	P1
$C_{T,opt}$	Offset for optimal photosynthesis temperature relationship	$^{\circ}\text{C}$	26.0/25.0	P1
$C_{T,min}$	Offset for minimal photosynthesis temperature relationship	$^{\circ}\text{C}$	–2.0/–4.0	P1
$C_{T,max}$	Offset for maximum photosynthesis temperature relationship	$^{\circ}\text{C}$	38.0	P1
SLA	Specific leaf area (LAI per dry matter content)	g m^{-2}	0.026/0.00926	P1
LAI_{MAX}	Maximum LAI	$\text{m}^2 \text{m}^{-2}$	5.0	P1
$K_{lai,happy}$	LAI threshold to stop using carbohydrates	–	0.5	P1
F_{stress}	Limit to photosynthesis from soil water stress	–	6.0	P1
Phenology				
$K_{pheno,crit}$	Multiplicative factor for growing season start threshold	–	1.0	P1
$C_{T,senescent}$	Offset for temperature threshold for senescence	$^{\circ}\text{C}$	12.0	P1
$L_{age,crit}$	Average critical age for leaves	days	180/910	P1
$\tau_{leafinit}$	Time to attain the initial foliage using carbohydrate reserve	days	10	P1
Soil water availability				
Hum_{cste}	Root profile	–	0.8/1.0	P1
Respiration				
Q_{10}	Temperature dependence of heterotrophic respiration	–	1.99	P1
K_{soilC}	Multiplicative factor of initial carbon pools	–	1.0	P1
$HR_{H,a}$	First-degree coefficient of the moisture control of heterotrophic respiration	–	–1.1	P1
$HR_{H,b}$	First-degree coefficient of the moisture control of heterotrophic respiration	–	2.4	P1
$HR_{H,c}$	Offset of the function for moisture control factor of heterotrophic respiration	–	–0.29	P1
$HR_{H,min}$	Minimum value of the function for moisture control factor of heterotrophic respiration	–	0.25	P1
MR_a	Slope of the relationship between temperature and autotrophic respiration	–	0.16	P1
MR_b	Offset of the relationship between temperature and autotrophic respiration	–	1.0	P1
GR_{frac}	Fraction of biomass available for growth respiration	–	0.28	P1
Z_{decomp}	Scaling depth determining effect of soil water on litter decomposition	–	0.2	P1
$H_{crit,litter}$	Scaling depth determining the litter humidity	–	0.08	P1
Energy balance				
$ZO_{overheight}$	Roughness length	m	0.0625	P1
$K_{albedo,veg}$	Multiplying factor for surface albedo	–	1.0	P1
Allocation				
r_0	Controlling allocation to fine roots with light and belowground limitations (Eq. (B1))	–	0.3	P2
s_0	Controlling allocation to sapwood with light and belowground limitations (Eq. (B2))	–	0.3	P2
s_{init}	Initial allocation of sapwood aboveground at $A_{ge} = 0$ (Eq. (1))	–	0.2	P2
d_{alloc}	Time constant for aboveground sapwood allocation defining the inertia of age-dependent allocation of sapwood to aboveground biomass (Eq. (1))	year	5	P2
T_{res}	Residence time of all biomass compartments. The inverse of this is mortality	year	40	P3

After reaching the minimum of the cost function, a first order measure of the parameter uncertainties and error correlations can be simply derived by assuming linearity of the model around the estimated parameters and Gaussian errors. The posterior error covariance matrix \mathbf{P}_a is calculated from the prior error matrices and the Jacobian of the model (i.e., \mathbf{H}_∞ , the derivative of all model outputs with respect to parameters at the minimum of the cost function) (Tarantola, 2005), as in the following equation:

$$\mathbf{P}_a = [\mathbf{H}_\infty^t \mathbf{R}^{-1} \mathbf{H}_\infty + \mathbf{P}_b^{-1}]^{-1}. \quad (4)$$

Valuable information can be derived from \mathbf{P}_a , such as how much constraint can be brought by the data on each parameter (diagonal of \mathbf{P}_a), and the level of a posteriori correlations between the parameters (non-diagonal terms that appear after the optimization, indicating for instance equifinality of parameters). The Jacobian (\mathbf{H}_∞) is calculated with the finite difference method. However, given that the parameter prior errors are not Gaussian but truncated Gaussian PDFs, the posterior PDFs are also truncated and the estimated uncertainties (i.e., the 1-sigma interval) will represent a larger interval than the standard Gaussian 68.27%. They should be used as a first order approximation to evaluate the relative constraint brought by the observations on the different parameters.

The observation error that should account for both model and measurement errors was defined as the RMSE between the observations and the model simulations using the a priori parameter values (Kuppel et al., 2012). This should include for example the uncertainties related to partitioning the net CO₂ flux into its component fluxes (GPP and TER). Micrometeorological flux measurements include both random (e.g. turbulence sampling error and instrument error) and systematic errors (e.g. instrument calibration and design and data processing) (Lasslop et al., 2010; Richardson et al., 2012). Estimation of biomass at stand level is subject to errors from sampling and measurement aggregation from individuals to the stand scale (Wutzler et al., 2008) as well as errors in allometric equations.

For the biomass observations only annual values were available. Therefore there is an imbalance in the number of observations that were available for the different data streams. A proper accounting of all error correlations in \mathbf{R} would be the most rigorous way to address this issue. Given that we were not able to characterize these off diagonal terms, the imbalance was partly compensated for by dividing each term of the cost function by the number of observations (division by N_i in Eq. (3)) to avoid a situation where the optimization solution is dominated by the assimilation of data-streams with denser sampling. This weighting approach removes possible aliasing of biases in rich data streams onto parameters related to sparse data streams (Wutzler and Carvalhais, 2014), and thus likely decreases the overall impact of model and observation biases. However, since the weighting is removing some of the certainty we have in the rich data streams, the posterior error on the parameters constrained by the rich data streams is likely overestimated (Wutzler and Carvalhais, 2014).

2.4. Optimization protocol

2.4.1. Optimized parameters

We selected parameters based on a preliminary sensitivity analysis and expert evaluation of the physical equations controlling the uptake, allocation and release of carbon in ORCHIDEE (see Kuppel et al., 2012). Since Le Bray is a coniferous forest a few parameters related to only deciduous leaf phenology were left out. In contrast to previous studies dedicated to the optimization of ORCHIDEE parameters (see the list of studies under: <https://orchidas.lscce.ipsl.fr/results.php>), we included five additional parameters related to the allocation of carbon in the different reservoirs because in this

study we are assimilating biomass data, as described above. The a priori range of all parameters was based on expert knowledge and the a priori uncertainty for each parameter was set to 40% of that range with further constraint provided by defined physical boundaries (shown in the Supplementary material in Table A1). The parameters and their a priori values are shown in Table 1 and the equations using these parameters are given in Appendix A. Parameter set P1 in Table 1 consists of parameters related to carbon and energy fluxes and parameter set P2 includes the four parameters related to allocation. Parameter set P3 consists of only one parameter, the living biomass carbon residence time.

2.4.2. Simulations and optimizations performed

We adopted the following approach for the model simulations and optimizations, in order to investigate the different objectives of our study:

2.4.2.1. Forward simulations. Three different protocols were used for the forward simulations in order to examine the difference between forest in steady state or a forest grown to be its realistic age.

- “SS”:** Spin-up of different carbon pool sizes followed by a transient simulation over periods 2001–2004 (Hesse) and 2001–2005 (Le Bray) with the forest at “equilibrium” (i.e., a forest in steady state whose age does not match that of the site)
- “RA”:** Spin-up of the different carbon pool sizes followed by a clear-cut of the forest and a similar transient simulation to match the real age of the forest during the period of measurements.
- “RA-LAI”:** Similar to “RA” with LAI described as a function of age (Eq. (2)) – i.e. the new LAI scheme.

The spin-up is performed by cycling of the present day climate to accumulate enough soil carbon so that the net (decadal-scale) carbon flux between ecosystem and the atmosphere is close to zero. In order to simulate the real age of the forest, biomass pools were all reset to zero, i.e. the biomass was removed from the system, after the spin-up and partly transferred to the woody litter pool, following the procedure developed by Bellassen et al. (2010). The forest was then allowed to grow to its true age driven by the meteorological forcing available at each site. The different model simulations performed in this study are summarized in Table 2.

2.4.2.2. Optimization scenarios. In the optimizations, however, only the forest with the realistic age was used. Given the lack of representation of forest management and natural disturbances in the model it was a challenge to use AGB_{tot} directly in the optimizations. In terms of biomass-related observations, measurements of AGB_{inc} more closely match what ORCHIDEE currently simulates and therefore we used these data as a starting point in our optimization. However, this is only a first step and our final aim is to be able to assimilate the total biomass stock and not the increment.

We decided to split the optimization process into two steps, the first step using the standard data stream set (flux data and AGB_{inc}, the annual increment of aboveground biomass) to optimize the fast processes and allocation in the model (“Step 1”). The two forests are mono-aged stands without natural regeneration by recruitment. The slow process of mortality is then optimized in a second step (“Step 2”) by adjusting the residence time parameter T_{res} . In order to optimize T_{res} the simulation period needs to correspond to the actual forest age since planting after the clearcut, as opposed to just the measurement time period, so that the measured total biomass value can be compared to the modeled one. This increases the computing time of the optimization procedure significantly (i.e. by several days), which hampers the use of a simultaneous approach with all parameters optimized at once including all data

Table 2
The short names and descriptions of the different simulation runs and optimizations.

Name of the run	Description
Initializations	
SS	Forest is in steady state.
RA	Clear cut is simulated for the forest after steady state and then it is simulated to be in its real age.
RA.LAI	Like RA with the new LAI scheme (Eq. (2)) included.
Optimizations	
REF_prior	The reference run before optimization.
REF_post	The reference run with data streams GPP, TER, ET and AGB.inc, after optimization using parameter sets P1 and P2.
T1.1	Optimization done with partitioned fluxes, GPP, TER and ET and parameter set P1.
T1.2	Optimization done with AGB.inc, using parameter set P2.
T2	Optimization done with AGB.tot, using parameter set P3.

streams. The advantage of such a stepwise approach is that we can use a short simulation period for the Step 1 optimization in order to optimize many (over 20) parameters for the “fast” processes and allocation, and then the required longer simulation to optimize just T_{res} in step 2. Optimizing with the total biomass in Step 2 will likely not have a strong feedback on the “fast parameter” values (having only influence on the magnitude of the autotrophic respiration which depends on biomass) and therefore there is no strong need to further iterate after Step 2.

In Step 1 we also performed different test cases by changing the number of data streams assimilated, as well as the number of parameters. This provided insight about the importance and role of the different data streams/parameters in order to answer the questions posed in the introduction section, e.g. how much the information content of biomass data adds to the optimization (see Tables 2 and 3). After each optimization we compare the prior and optimized model performance and evaluate the impact of the different data streams on the posterior parameters.

The different optimizations performed in Step 1 and Step 2 are summarized in Table 3. Additionally we made future climate simulations with the original parameter set and the ones obtained from the optimization. These simulations and their results are described in Supplementary Material, section B.

3. Results and discussion

3.1. Steady state versus real age forest biomass in forward simulations

Aboveground biomass (AGB.tot) simulated with the steady state assumption (SS) (Section 2.5.2. 1a) was highly overestimated for both Hesse and Le Bray, which were 36 and 31 years old in 2001, respectively (see Fig. 2). The ORCHIDEE simulations with the standard parameters overestimated AGB.tot at Hesse by 223.6% and by 136.1% at Le Bray over the period 1997–2005. When simulating the real forest age (RA) (Section 2.5.2 1b) the simulated biomass was significantly lower, and the agreement with data was improved. After this change the AGB.tot remained overestimated by 65.0% at Hesse however and underestimated by 11.9% at Le Bray.

Table 3
The abbreviations for different test runs with used data streams, optimized parameter sets and motivation for each test.

Test	Used data streams	Optimized parameters	Motivation
REF	GPP, TER, ET, AGB.inc	P1, P2	The reference optimization with partitioned fluxes and biomass increment. The inversion makes use of maximum available information on daily to inter-annual time-scale processes.
T1.1	GPP, TER, ET	P1	To assess the role of the AGB.inc by comparing to REF. P2 parameters were not included as their influence on the three fluxes is rather weak.
T1.2	AGB.inc	P2	To assess how well the biomass data alone can constrain the allocation related parameters.
T2	AGB.tot	P3	Optimization in a second step using the optimized parameters from the “REF” optimization. The aim is to demonstrate how to best use the AGB.tot data in the optimization in combination with other data.

The overestimation of simulated aboveground biomass at Hesse was caused by several factors. Disturbances that resulted in decreases in the biomass, such as thinnings (which took place at Hesse in 1999 and 2005, Fig. 2a) have not been taken into account in the current version of the model.

A second factor contributing to the large discrepancy in the modelled biomass was related to the fact that the ORCHIDEE AR5 model does not describe the early stages of the forest growth in realistic way. When the new LAI scheme was used (“RA.LAI” run) the positive bias of AGB.tot was reduced to only 11.3% for the Hesse forest. However, for Le Bray the fit to the AGB.tot observation was degraded with the “RA.LAI” simulation, increasing the underestimation of AGB.tot to 24.8%, which indicated that the carbon input to the aboveground woody biomass is too low. Following this initial evaluation we decided to use the new LAI scheme for Hesse only.

The success of the new LAI scheme at Hesse but not at Le Bray brings out another issue related to large-scale modelling. The new LAI scheme has been parameterized using permanent monitoring plots, yield tables and French inventory data, but even though it is very successful for some forest ecosystems in France, the biomass is underestimated for coniferous forests in southwestern France without optimization (Bellassen et al., 2011). The maritime pine at Le Bray is a fast growing species and the proposed generic dependence of maximum LAI on forest age (Eq. (2)) is probably not well adapted. To overcome these issues, we could increase the number of species in the model with specific LAI dependence equations, instead of relying on generic PFT classifications, or move toward the use of a more generic carbon allocation scheme that does not require the prescription of a maximum LAI (ORCHIDEE-CAN recent version described in Naudts et al., 2015).

3.2. Optimization with eddy covariance fluxes and biomass increment

3.2.1. Results from the optimization of GPP, TER, ET and AGB.inc (REF)

3.2.1.1. Hesse. Daily GPP, TER and ET measurements were used in the REF optimization, the Step 1 optimization, simultaneously with annual aboveground biomass increment at Hesse for years 2001–2004. The results are shown in Fig. 3. There was a

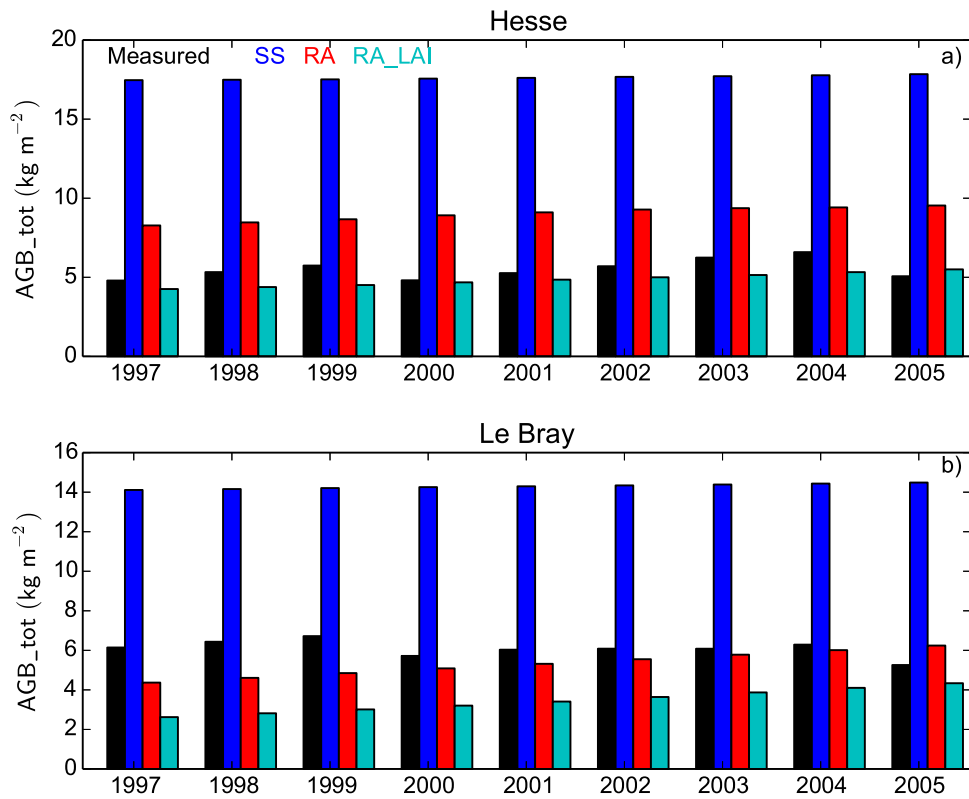


Fig. 2. The measured aboveground biomass at Hesse and Le Bray in 1997–2005 displayed in black bars, the simulated aboveground biomass at steady state (SS) in blue and the simulated aboveground biomass in realistic age (RA) in red. In cyan is shown the simulated aboveground biomass using the new LAI scheme (RA.LAI).

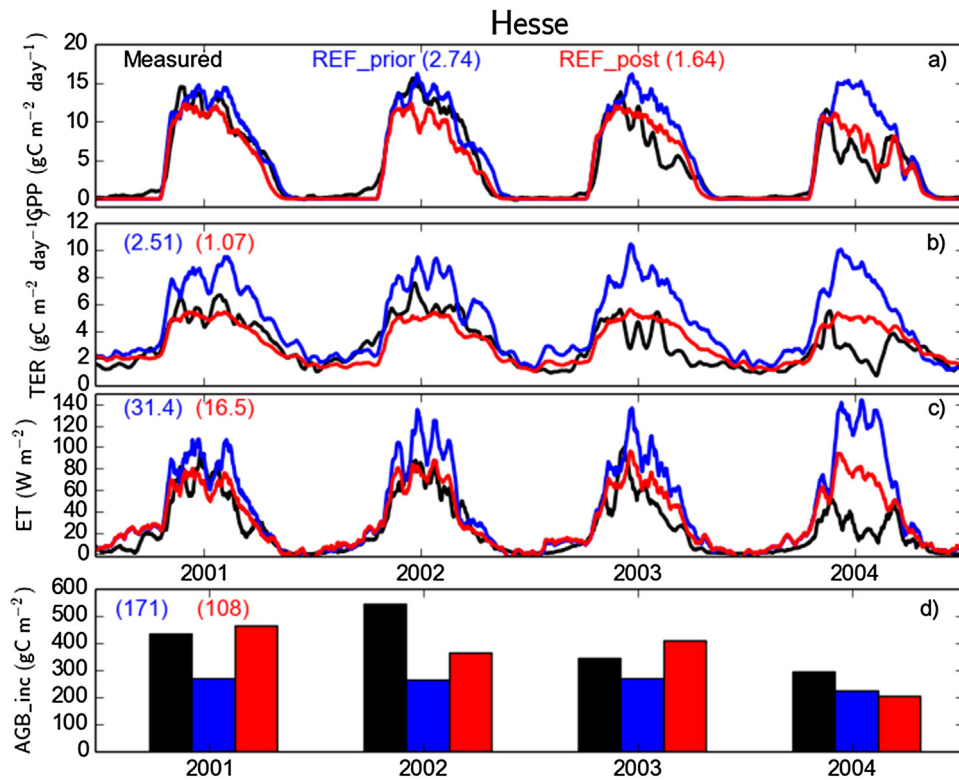


Fig. 3. The micrometeorological fluxes (GPP, TER and ET) and annual aboveground biomass increments at Hesse in 2001–2004. The 15-day running averages of the flux daily values are shown. The measurements are shown in black, a priori simulation result as blue and a posteriori in red. The RMSE value between simulation and observation is shown in parenthesis.

Table 4
RMSE values and biases (in parenthesis) for different observations between various optimizations at Hesse and Le Bray.

	GPP [$\text{g C m}^{-2} \text{ day}^{-1}$]	TER [$\text{g C m}^{-2} \text{ day}^{-1}$]	ET [W m^{-2}]	AGB.inc [g C m^{-2}]
Hesse				
Prior	2.74 (1.0)	2.51 (1.7)	31.40 (17)	171.3 (–149)
REF_post	1.64 (–0.28)	1.08 (0.13)	16.53 (7.7)	108.2 (–44)
T1.1	1.76 (–0.49)	1.03 (0.07)	12.76 (2.8)	193.7 (–172)
T1.2	2.64 (0.90)	2.18 (1.40)	30.59 (16.7)	87.1 (15)
Le Bray				
Prior	1.72 (–0.31)	1.47 (1.01)	19.10 (–4.4)	51.3 (–49)
REF_post	1.65 (0.08)	1.01 (–0.02)	15.72 (–2.1)	15.8 (6)
T1.1	1.63 (–0.03)	1.04 (0.04)	15.83 (–1.3)	73.7 (71)
T1.2	1.77 (–0.44)	1.31 (0.80)	19.43 (–4.9)	11.9 (–5)

RMSE values between observations and model estimates between years 2001–2004 at Hesse, 2001–2005 at Le Bray. The time period 2001–2004 was used in the optimizations at Hesse, and 2001–2005 at Le Bray. The prior values are the default values in ORCHIDE before optimization. The reference optimization refers to optimization using simultaneously GPP, TER, ET and AGB.inc. The last two lines refer to optimizations where only GPP+TER+ET, or AGB.inc were used, respectively (see Table 3). The units for each variable are shown in brackets.

considerable reduction in the RMSE between the model and the observations for all four data streams after the optimization (see Table 4, columns ‘REF.prior’ and ‘REF.post’).

At Hesse there was a drought during spring and summer 2003 and spring 2004 (124 and 100 days with soil water deficit, respectively; Granier et al., 2008) compared to 2001 and 2002 (23 and zero days with soil water deficit), which were considered “normal” years. This inter-annual variability affected the observed carbon and water exchanges between the vegetation and atmosphere at the site (Granier et al., 2008), with GPP and TER being strongly correlated with soil water content and even premature leaf fall reported in 2003 (Bréda et al., 2006). The a priori simulated GPP did not replicate the observed decrease in uptake by the vegetation during the two dry years, and the annual mean GPP was overestimated by 41.9% instead of by 6.6% only in 2001–2002 (Fig. 3a). The a posteriori simulated GPP was able to capture the decreased uptake caused by drought, yet with a residual overestimation of the annual mean optimized GPP of 14.3% and 10.2% for 2003 and 2004, respectively. However, after the optimization carbon uptake in year 2002 became underestimated (by 23.3%), whereas the a priori carbon uptake was accurate (i.e., 1.7% from observed). The MDF did not alter the seasonal dynamics of the simulated GPP greatly, but diminished the CO_2 uptake in autumn, anyhow not leading to underestimates in the posteriori results. The simulated GPP also showed a slightly more abrupt increase than observed during spring. For beech forests, after the leaf expansion, there is a gradual increase of photosynthetic capacity in the leaves, and this is not included in the model.

The a priori simulated annual mean TER was overestimated by 54.5% (Fig. 3b). Although the simulated TER reached the observed magnitude after optimization (within 0.2% for all four years), the optimized model results did not fully replicate the magnitude of the observed decline in respiration caused by the drought. Yet, the overestimation of TER for 2003–2004 was 92.6% before the optimization and was reduced to 23.4% after optimization. This moderate inability of the model to capture TER reduction in dry periods, even after optimization, may indicate limits of the formulation of drought response of the decomposition processes in the soil in ORCHIDE, or limits of the linear dependence of autotrophic respiration to GPP. However, there is uncertainty in the TER estimate in the flux separation (Raj et al., 2016), and it is not known how well the separation method works during drought conditions, and therefore the performance of the optimized model might be more satisfactory than revealed in Fig. 3b.

The a priori simulated evapotranspiration was overestimated in all cases, with the summer time (June–August) average ET overestimated by 52.5% and a RMSE of 31.4 W m^{-2} (Fig. 3c). After optimization, this overestimate and the RMSE dropped to 19.6% and

16.5 W m^{-2} , respectively, thus improving the magnitude of the ET. The decline in the observed ET magnitude during the dry summers was reproduced in the MDF results, but the temporal dynamics of the drought response were not well captured. This is likely connected to the fact that the model is missing a process representing increased soil–root resistance to water flow when soil water potential is very low or a mismatch between simulated and real soil water contents. A newer version of ORCHIDE has included a fully mechanistic scheme for the plant hydraulic architecture (Naudts et al., 2015), thus issues such as this will be investigated again in future studies using this version of the model.

The a priori simulations underestimated the magnitude of the AGB.inc by 148.8 g C m^{-2} (Fig. 3d). The a priori simulation also had a much lower annual AGB.inc variability than observed, with a standard deviation of 19.5 g C m^{-2} compared to 95.1 g C m^{-2} in the observations. After MDF, the optimized mean AGB.inc was in better agreement with the observation, with a negative bias of 44.3 g C m^{-2} ; however, we did not improve the correlation of the interannual variation between the observations and the model, even though the standard deviation became comparable to the observations. The measured AGB.inc was the highest in 2002 and dropped in 2003 and 2004, but this behavior was not seen in the modeled values.

3.2.1.2. Le Bray. Similar to Hesse, the optimization improved the model performance for the four variables at Le Bray (Fig. 4 and Table 4). Overall, the prior model data fit was better at Le Bray than at Hesse and the performance of the optimized model in terms of RMSE was relatively similar between the two sites. The model was not able to capture the interannual dynamics of GPP even after the optimization and the summertime magnitude of GPP remained lower than that of the observations in years 2001, 2003 and 2004 (Fig. 4a). Before optimization the model overestimated the average annual GPP by 9.2%, which was reduced by the optimization to 5.2%. After MDF the summertime TER levels matched the observations better (the average annual overestimation being 18.5% before optimization, and only 3.5% after), but wintertime respiration during the last two years was still overestimated by the model (Fig. 4b). MDF did not change the ET much (Fig. 4c), with the summertime ET being underestimated before the optimization by 11.5% and after by 9.7%. The simulated annual AGB.inc at Le Bray was closer to the observations after MDF; the simulated average value was 49.5 g C m^{-2} higher than the observations before the optimization, but this positive bias was reduced to only 6.3 g C m^{-2} after optimization. However, even after the optimization, the year-to-year variations did not match that of the observations (Fig. 4d) like for the Hesse site, although the magnitude of inter-annual variation was closer to that of the observations. The standard deviation of the

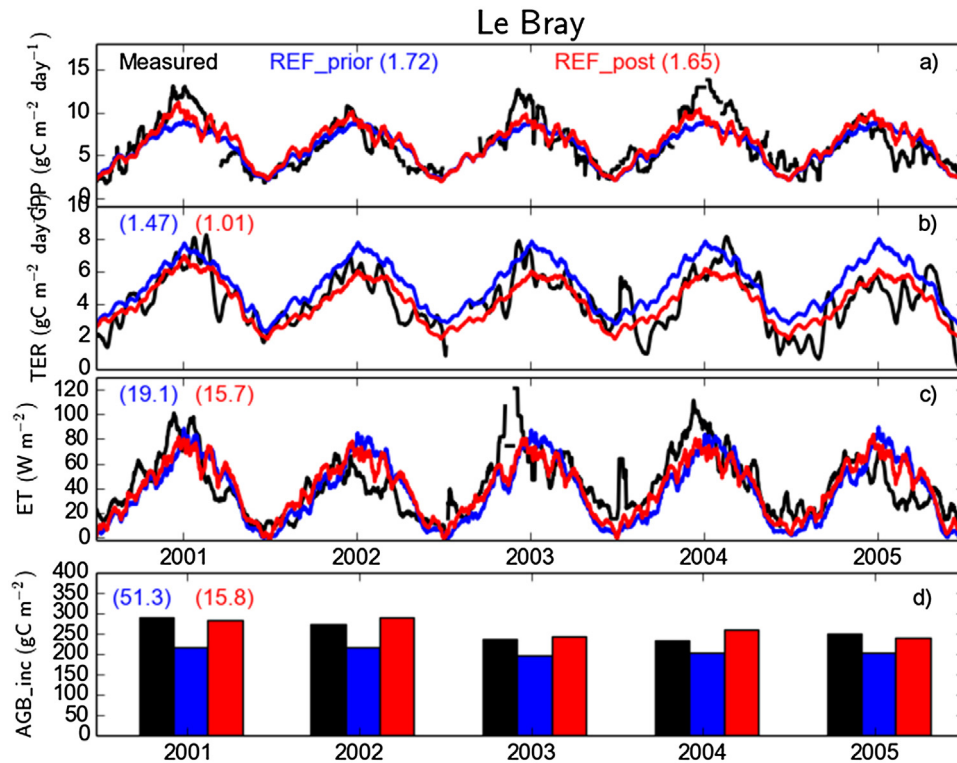


Fig. 4. The micrometeorological fluxes (GPP, TER and ET) and annual aboveground biomass increments at Le Bray in 2001–2006. The 15-day running averages of the flux daily values are shown. The measurements are shown in black, a priori simulation result as blue and a posteriori in red.

observations was 21.4 gC m^{-2} . By comparison the model simulated an a priori standard deviation of only 9.0 gC m^{-2} , which increased to 18.6 gC m^{-2} after the optimization.

The optimization converged after 40 iterations at Hesse and after 100 iterations at Le Bray of the genetic algorithm, implying that the MDF method had more difficulties in finding the global minimum at Le Bray. The cost function values before and after the optimization as well as the reduced χ^2 values are shown in Table C1 in the Supplementary material. The cost function values differ between the cases due to different data streams assimilated and different parameters optimized. The lower reduced χ^2 values (after optimization) for the flux data at Hesse for the REF and T1.1 optimizations indicate better model performances at this site compared to Le Bray (smaller misfit). Reversely, the significantly lower values for the aboveground biomass increments at Le Bray for the REF and T1.2 cases indicate that the carbon allocation scheme better captures the dynamics of the aboveground growth of the pine forest site.

3.2.2. Trade-offs and co-benefits of assimilating each data stream

3.2.2.1. GPP+TER+ET (“T1.1”). In this test (T1.1) we omitted the AGB.inc from the reference case and only included the eddy covariance fluxes in the optimization, to investigate, by comparison with the reference optimization, the benefit of assimilating aboveground biomass increment in addition to the fluxes and to see if it modifies the resultant gross carbon fluxes. Not surprisingly, the largest difference was seen in the estimation of AGB.inc, which was worse at both sites when this variable was not assimilated and only the fast parameters are optimized (Table 4). In fact the fit to AGB.inc data after optimization of fluxes resulted in an even larger RMSE than the a priori value (Table 4), thus the fluxes degraded the fit to the data for AGB.inc. Otherwise, there were no major changes in the optimization performance compared to the REF case regarding the match to 15-day average eddy covariance fluxes, except the a posteriori RMSE for GPP was slightly worse than in the reference

case at Hesse, therefore the addition of AGB.inc (and the parameter set P2) also improved the flux optimization in the reference case.

3.2.2.2. AGB.inc (“T1.2”). The objective of this test was to assess the potential of using only AGB.inc, obviously a “weak” constraint on all the “fast” model parameters. Optimization of AGB.inc brings information about sapwood allocation, because although AGB.inc is determined by both GPP and TER processes, it is calculated roughly as a fraction of GPP minus the autotrophic respiration and allocation to other biomass compartments. Therefore, assimilating AGB.inc could potentially reinforce the constraints on GPP and TER.

As expected, the optimized AGB.inc was improved at both sites but the trade-off was that fit to the eddy covariance fluxes (that were not included in the optimization) deteriorated (Table 4). With respect to matching observed fluxes, this test was thus worse than the T1.1 case or the a priori results. This was not surprising as the AGB.inc data are more directly related to processes that are acting on an annual time step, and therefore do not contain enough information to constrain the “fast” surface fluxes operating at a half-hourly to seasonal timescales. Additionally, only a restricted set of parameters was used that are controlling only the carbon allocation. However, the reference case gave good RMSE for also AGB.inc, even though not as good as with this test; it is thus better to include the fluxes with AGB.inc and obtain the best fit also for GPP.

3.3. Optimization using total aboveground biomass in a second step (“T2”)

As described in Section 2.5.2 we only optimize T_{res} using total aboveground biomass in the second step. Fig. 5 compares the observed total aboveground biomass together with the a priori simulated biomass and the simulated biomass after Step 1 (results described in Section 3.2.1) and after Step 2. The simulated biomass from the REF inversion at both sites (Step 1) showed a large

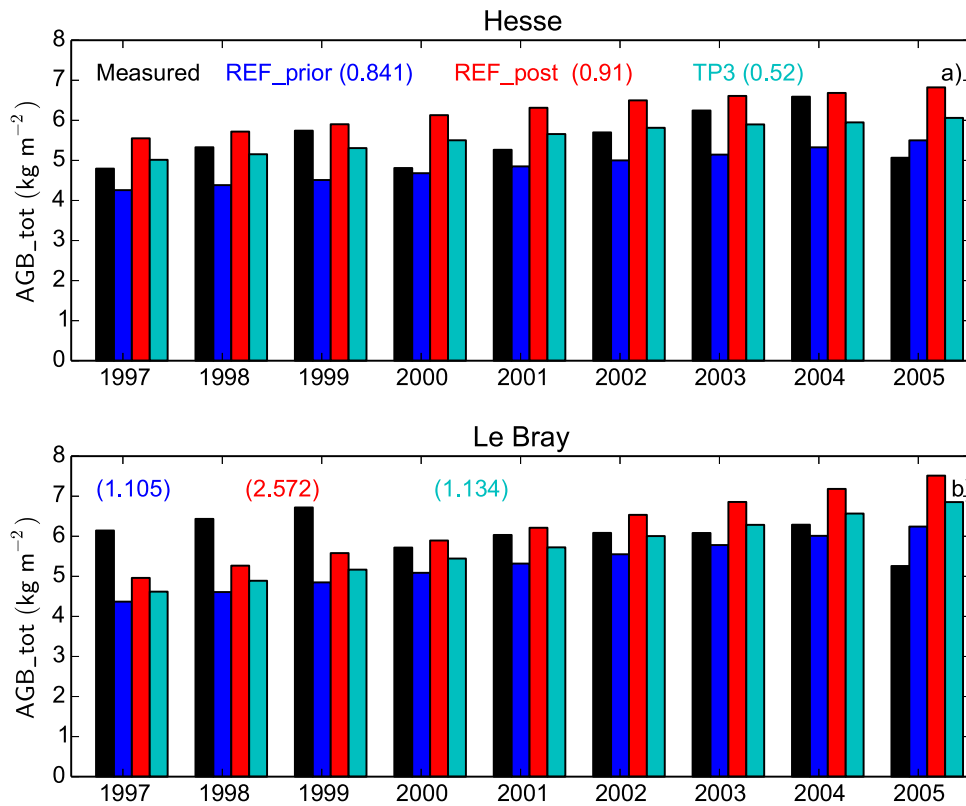


Fig. 5. The aboveground biomass at Hesse (a) and Le Bray (b). The observed aboveground biomass is shown in black, the prior model estimate in blue (REF.prior), the model estimate in “Step 1” (REF.post) in red and the model estimate in “Step 2” (T2) in cyan. The RMSE values between the observation and model estimates are shown in parenthesis in units kg C m^{-2} .

discrepancy with observed AGB_{tot} , since the use of AGB_{inc} in Step 1 led to an increase in the annual biomass production from the prior at both sites. This was because the total biomass observations showed a decrease in some years due to disturbance or management which was not accounted for in the biomass increment.

After optimization of the residence time in Step 2 the simulated AGB_{tot} was found to be much closer to the observations at Hesse, with an RMSE of 0.52 kg C m^{-2} and overestimation of just 1.7%. This was achieved as a result of a large decrease in the residence time from its default value of 40 years down to 16.8 years (i.e. factor of two increase in mortality). This is a quite low residence time for woody matter, likely explained by the fact that in our set-up, the adjustment of T_{res} implicitly accounts for non-modelled biomass carbon removal by disturbances (mainly thinning events, plus additional mortality related to climate extremes such as the 2003 drought) that are taken into account in the observations (as mentioned above).

When a simulation was performed with both optimized residence time and optimized parameters from the REF.post, the new value of T_{res} did not have a strong impact on the GPP, TER and ET flux simulations except for a small decline in the magnitude of TER. This was due to a lower autotrophic respiration because of the smaller amount of biomass. However, the new residence time had a strong impact on the AGB_{inc} , resulting in lower values for this variable due to the faster turnover rate causing more biomass to go to the litter pool. The annual increment with optimized T_{res} became so small that it actually was negative in 2004 (see Supplementary Material, section D), when the GPP was the lowest for the whole period. Therefore, the short residence time obtained by optimizing the model to fit the total AGB_{tot} becomes unrealistic (when interpreted as a constant mean mortality) and it falsified one component of the model (AGB_{inc}).

The optimized residence time at Le Bray was 31 years, which was again lower than the a priori value of 40 years. However, the posterior RMSE value for AGB_{tot} at Le Bray in Step 2 was slightly higher ($1.134 \text{ kg C m}^{-2}$) than the a priori RMSE ($1.105 \text{ kg C m}^{-2}$). This result might be surprising, but it just reveals that changing the “fast” parameters in Step1 (P1 and P2 sets) does not allow to fit the AGB_{tot} as well as with the prior values and also the inclusion of AGB_{inc} changes allocation parameters without taking into account that the total biomass might also decrease. Note that the simulated biomass linearly increased over time while the observations showed a more stable evolution with large changes in 1999–2000 and 2004–2005. Similarly to Hesse, the new optimized residence time did not influence the model performance for the fluxes, but it did degrade the AGB_{inc} estimate as was seen at the Hesse site, even though it was not negative in any year like at Hesse.

The information contained in AGB_{tot} relates to processes that operate during the whole life-time of the forest prior to observation. AGB_{tot} assimilation has a negligible influence on CO_2 fluxes in the model as such (only a weak effect through maintenance respiration), but a larger effect on carbon storage, so this type of measurements can thus be used to constrain longer term processes such as mortality and management (if not explicitly taken into account in the model). Large data sets exist for biomass, e.g. from remote sensing and forest inventories, that have much better spatial coverage than sparse network of flux measurements. However these datasets do not provide the age of the forest, which thus further complicates the assimilation of these observations. Nevertheless in upscaling studies the fast parameters could be constrained with a few flux towers and more numerous biomass observations could be used to constrain the slow parameters, though as mentioned we need ancillary information of the forest age.

3.4. Evaluation of the optimized model

3.4.1. Optimized parameters and error estimation

The errors of the parameters were greatly reduced from their prior values after optimization at both sites, and for all the different test cases that were performed with different data streams (i.e. REF compared to T1.1 and T1.2 Table 5). In addition to the optimized parameter value and uncertainties, the cross-correlations between the different parameters reveal important information. If the parameters drift to the margins of their prior range, this would indicate problems with sub-model structures. However, this did not occur at either site.

At Hesse, flux related parameters (P1) differed between each test and it is not straightforward to evaluate the differences between each case due to parameter correlations. The V_{cmax} parameter decreased in the reference case and increased in the test case T1.1 when AGB data were not used (Table 5), likely showing that optimizing allocation-related parameters has an influence on parameters controlling photosynthesis. Leaf level observations of V_{cmax} show a large spread of values (Montpied et al., 2009) and the values from the optimizations are within this range. Observations at another beech forest (Epron et al., 1999) and a model inversion study (Verbeeck et al., 2008) suggest smaller values of V_{cmax} (than the default value of ORCHIDEE) are more appropriate for temperate deciduous forests, implicating that the REF case gave the most reasonable result in this respect. The LAI_{MAX} parameter, which defines a peak LAI value that cannot be surpassed in the model, decreased too much ($4.3 \text{ m}^2 \text{ m}^{-2}$) when compared to the observed maximum LAI reached in the peak growing season ($7.6 \text{ m}^2 \text{ m}^{-2}$).

The allocation-related parameter r_0 , which controls the allocation to fine roots (Eq. (B1)), decreased in the REF and T1.2 cases in which it was optimized, whereas the parameter s_0 controlling allocation to sapwood (Eq. (B2)), increased in both optimizations. d_{alloc} (Eq. (1)), which is the time constant for the aboveground sapwood allocation, also moved in the same direction (decreased) in both cases. However, the S_{init} parameter (Eq. (1)), which defines the initial allocation of sapwood derived from NPP to aboveground compartment, decreased strongly in the reference case but increased in the T1.2 test. The differences between the REF and T1.2 cases were caused by the fact that in the REF case the fluxes and AGB.inc were constraining other parameters that exerted an influence over S_{init} . In the T1.2 case (AGB.inc only) many large correlations between the allocation related parameters exist, the largest being positive correlation between r_0 and S_{init} (0.43). This nicely shows that the eddy covariance fluxes in the REF case bring extra information that further resolves the parameter space compared to an optimization with biomass data alone.

At Le Bray, we obtained similar parameter error correlations and similar complex patterns between the REF, T1.1 and T1.2 cases. The optimized V_{cmax} increased in both cases (REF and TP1.1), which is in line with observations (Medlyn et al., 2002) and the previous MDF study (Santaren et al., 2007). At Le Bray the LAI_{MAX} value increased in both cases, which was in the opposite direction to the observed LAI. The allocation parameters r_0 and s_0 increased in both the reference and T1.2 cases.

3.4.2. Evaluation against other in situ data

The results from the simulations using the optimized parameter set from the REF inversion at Hesse were compared to a data set from Granier et al. (2008) having different measurements at the site. We first considered the fine root production as calculated from the turnover rate, as only the fine root production was observed by Granier et al. (2008) and not the maximum size of fine root carbon pool. The fine root production of the a priori model was much higher than the observations, with an annual average of 218.4 g C m^{-2} (standard deviation of $\pm 7.0 \text{ g C m}^{-2}$) compared to

the observed value of 68.2 g C m^{-2} ($\pm 9.7 \text{ g C m}^{-2}$). This positive bias was reduced after the optimization, with an optimized annual average of 41.7 g C m^{-2} ($\pm 29.1 \text{ g C m}^{-2}$). This result is consistent with the adjustment of parameter r_0 , which controls allocation to fine roots, and was reduced after optimization (0.165 instead of the a priori value of 0.300; see Table 4).

The a priori model also overestimated leaf production with an annual average of 214.1 g C m^{-2} ($\pm 7.0 \text{ g C m}^{-2}$), compared to the observed value of 157.8 g C m^{-2} ($\pm 22.8 \text{ g C m}^{-2}$). The a posteriori simulation resulted in a lower annual average of 126.2 g C m^{-2} ($\pm 3.2 \text{ g C m}^{-2}$) that is closer to the observations. As with the fine root production, the leaf production has been calculated from the turnover rate in order to compare the model output with the measurements. In summary, the allocation to both fine roots and leaves was reduced after the optimization. This is consistent with the fact that more biomass was allocated to sapwood after optimization, as seen in Fig. 4d, and also in the increase of parameter s_0 (from 0.30 to 0.42), which controls the allocation of assimilated carbon to sapwood. That the allocation in the optimized model is therefore more consistent with the observations after MDF, highlights the added benefit of using annual aboveground biomass increment data in the optimizations. However, simulations with optimized parameters were still not able to capture the inter-annual variation seen in the observations of fine root and leaf production.

The observed autotrophic respiration (maintenance plus growth) was estimated to be 72.2% of the total ecosystem respiration, based on Granier et al. (2000). The observed value has significant uncertainty due to the partitioning and the presence of nighttime advection at the site (Aubinet et al., 2005). The a priori model overestimated autotrophic respiration – the annual average was $974.2 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 44.6 \text{ g C m}^{-2} \text{ year}^{-1}$), while the observed value was $736.2 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 97.7 \text{ g C m}^{-2} \text{ year}^{-1}$). After optimization, the simulated annual average for the autotrophic respiration was reduced to $528.3 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 19.3 \text{ g C m}^{-2} \text{ year}^{-1}$). The ratio of autotrophic respiration to TER was 52.5% before the optimization and 37.7% after, much lower than in Granier et al. (2008), demonstrating that although the decrease in TER after optimization resulted in a better fit to the data overall, the autotrophic respiration was reduced too much compared to heterotrophic respiration.

The observed annual average of NPP was $690.6 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 147.9 \text{ g C m}^{-2} \text{ year}^{-1}$) (Granier et al., 2008). Both the a priori and a posteriori model results overestimated NPP by almost a factor of two, with an a priori value of $1167.5 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 50.7 \text{ g C m}^{-2} \text{ year}^{-1}$) and an a posteriori value of $1148.2 \text{ g C m}^{-2} \text{ year}^{-1}$ ($\pm 89.3 \text{ g C m}^{-2} \text{ year}^{-1}$) – thus almost no change after the optimization. This was partly due to the fact that even though the optimizations resulted in a lower GPP, the autotrophic respiration was also decreased, and therefore the NPP remained too high. The optimized simulations were also not able to capture the strong inter-annual variability seen in the observations of NPP and autotrophic respiration, as for the fine root and leaf production.

4. Perspectives on assimilating biomass data into global terrestrial ecosystem models

Including AGB.inc in an optimization together with the flux data enabled the optimization of the allocation-related parameters without any deterioration compared to only using flux data in the optimization. However, although the new LAI scheme did help to improve the simulated a priori AGB.tot with respect to the observations at Hesse, we faced some challenges when assimilating both aboveground increment and total aboveground biomass within the ORCHIDEE model structure and therefore used a

Table 5
The parameter values and their uncertainties (in parenthesis).

Hesse	Prior	REF_post	T1.1	T1.2
P1 parameters				
$V_{c(max)}$	55.0 (22.4)	51.7 (4.64)	67.2 (2.93)	–
$G_{S,slope}$	9.00 (2.40)	6.67 (0.407)	6.55 (0.566)	–
$C_{T,opt}$	26.0 (6.40)	18.3 (1.88)	19.5 (0.832)	–
$C_{T,min}$	–2.00 (4.00)	0.871 (0.524)	–2.37 (0.218)	–
$C_{T,max}$	38.0 (16.0)	54.8 (8.89)	22.9 (0.348)	–
SLA	0.0260 (0.0148)	0.0388 (0.00549)	0.0412 (0.00155)	–
LAI_{MAX}	5.00 (2.00)	4.28 (0.237)	3.40 (0.116)	–
$K_{lai,happy}$	0.500 (0.140)	0.540 (0.0192)	0.500 (0.0264)	–
F_{stress}	6.00 (3.68)	2.64 (0.751)	2.21 (0.411)	–
$K_{pheno,crit}$	1.00 (0.480)	1.04 (0.0158)	1.00 (0.480)	–
$C_{T,senescence}$	12.0 (8.00)	13.5 (0.112)	12.0 (0.144)	–
$L_{age,crit}$	180 (48.0)	165 (7.27)	172 (7.54)	–
$\tau_{leafinit}$	10.0 (10.0)	12.0 (3.05)	10.0 (3.16)	–
Hum_{cste}	0.800 (1.12)	2.91 (0.498)	1.77 (0.608)	–
Q_{10}	1.99 (0.600)	1.77 (0.153)	1.39 (0.157)	–
K_{soilC}	1.00 (1.50)	0.287 (0.231)	0.358 (0.434)	–
$HR_{H,a}$	–1.10 (0.800)	–1.72 (0.800)	–1.53 (0.800)	–
$HR_{H,b}$	2.40 (1.68)	5.27 (1.67)	1.89 (1.68)	–
$HR_{H,c}$	–0.290 (0.600)	–0.388 (0.590)	–0.257 (0.600)	–
$HR_{H,min}$	0.250 (0.200)	0.553 (0.200)	0.410 (0.162)	–
MR_a	0.160 (0.0640)	0.101 (0.0362)	0.0938 (0.0296)	–
MR_b	1.00 (0.760)	0.763 (0.576)	0.814 (0.366)	–
GR_{frac}	0.280 (0.0640)	0.201 (0.0255)	0.280 (0.0259)	–
Z_{decomp}	0.200 (1.98)	3.19 (1.97)	2.48 (1.97)	–
H_{crit_litter}	0.0800 (0.196)	0.215 (0.180)	0.377 (0.196)	–
$ZO_{overheight}$	0.0625 (0.0320)	0.0382 (0.00820)	0.0625 (0.00714)	–
$K_{albedo,veg}$	1.00 (0.160)	1.13 (0.120)	0.988 (0.158)	–
P2 parameters				
r_0	0.300 (0.240)	0.165 (0.0880)	–	0.232 (0.107)
s_0	0.300 (0.240)	0.420 (0.0242)	–	0.461 (0.0659)
s_{init}	0.200 (0.160)	0.113 (0.0942)	–	0.223 (0.150)
d_{alloc}	5.00 (7.60)	2.36 (7.52)	–	3.80 (7.39)
P3 parameter				
T_{res}	40(35)	16.8 (1.08)	–	–
Le Bray				
	Prior	REF_post	P1	P2
P1 parameters				
$V_{c(max)}$	35.0 (12.0)	45.6 (3.26)	45.3 (1.17)	–
$G_{S,slope}$	9.00 (2.40)	9.40 (0.248)	10.4 (0.307)	–
$C_{T,opt}$	25.0 (6.40)	24.6 (0.992)	23.8 (0.426)	–
$C_{T,min}$	–4.00 (4.00)	–2.86 (0.599)	–2.93 (0.110)	–
$C_{T,max}$	38.0 (16.0)	24.9 (0.204)	25.2 (0.228)	–
SLA	0.00926 (6.4e ^{–3})	0.0199 (1.93e ^{–3})	0.0199 (3.72e ^{–3})	–
LAI_{MAX}	5.00 (1.60)	5.97 (0.358)	5.31 (0.0140)	–
F_{stress}	6.00 (3.68)	7.95 (2.78)	7.94 (2.15)	–
$L_{age,crit}$	910 (160.)	749 (29.6)	713 (27.4)	–
Hum_{cste}	1.00 (1.50)	2.68 (0.616)	2.58 (0.498)	–
Q_{10}	1.99 (0.600)	2.10 (0.158)	2.49 (0.225)	–
K_{soilC}	1.00 (1.50)	1.98 (0.220)	0.961 (0.275)	–
$HR_{H,a}$	–1.10 (0.800)	–0.151 (0.796)	–0.820 (0.788)	–
$HR_{H,b}$	2.40 (1.68)	4.21 (1.42)	2.80 (1.28)	–
$HR_{H,c}$	–0.290 (0.600)	–0.692 (0.515)	–0.311 (0.527)	–
$HR_{H,min}$	0.250 (0.200)	0.250 (0.164)	0.291 (0.174)	–
MR_a	0.160 (0.0640)	0.0924 (0.0379)	0.0818 (0.0333)	–
MR_b	1.00 (0.760)	0.188 (0.545)	0.580 (0.118)	–
GR_{frac}	0.280 (0.0640)	0.233 (0.0329)	0.288 (0.0212)	–
Z_{decomp}	0.200 (1.98)	0.190 (0.121)	0.503 (0.415)	–
H_{crit_litter}	0.0800 (0.196)	0.0288 (0.0150)	0.0610 (0.0730)	–
$ZO_{overheight}$	0.0625 (0.0320)	0.0997 (0.00238)	0.0994 (0.00181)	–
$K_{albedo,veg}$	1.00 (0.160)	0.812 (0.137)	0.860 (0.154)	–
P2 parameters				
r_0	0.300 (0.240)	0.488 (0.240)	–	0.319 (0.206)
s_0	0.300 (0.240)	0.490 (0.0315)	–	0.346 (0.0108)
s_{init}	0.200 (0.160)	0.00447 (0.0778)	–	0.200 (0.103)
d_{alloc}	5.00 (7.60)	18.5 (6.73)	–	2.90 (6.73)
P3 parameter				
T_{res}	40 (35)	31.2 (5.05)	–	–

stepwise approach to deal with this. The difference between these two sets of observations and the model is important to keep in mind for model evaluation and optimization. In the ORCHIDEE “AR5” version used in this study, in which mortality is a constant fraction of biomass, the change of AGB_tot and AGB_inc are always positive in a young growing forest not at equilibrium. In reality AGB_tot data might decline during a high mortality year due to disturbance or forest management. Therefore there is an inconsistency between these two observations and how the model is simulating the forest stand dynamics. Processes relating to disturbance and management (e.g. episodic canopy thinning) are not represented in this version of the ORCHIDEE model, and therefore using AGB_tot to optimize the long-term mean residence time resulted in an unrealistically low value. As a result the model simulated a negative aboveground biomass increment was modelled for one year, which was therefore opposite to the positive values seen in the observations. These results therefore imply that long-term residence time or turnover rate parameters cannot be optimized using total aboveground biomass data unless the history of disturbance and management are accounted for in the model.

A new version of ORCHIDEE that includes processes linked to stand growth and mortality, as well as management (including thinning; Naudts et al., 2015, based on Bellassen et al., 2010), will be a more appropriate model in terms of using biomass data in an assimilation framework. However, this version still does not describe mortality peaks due to natural disturbance and may still not capture year to year variation of AGB_tot. If the model did have an adequate representation of long-term forest dynamics, AGB_inc observations could be used to constrain modelled woody biomass increment of surviving trees, i.e. the annual NPP increment, in order to calibrate parameters relating to the natural fluxes and allocation. The AGB_tot measurements could then be used to optimize the trajectory of the total aboveground biomass, including parameters related to disturbance and human-related management as discussed above. Whether or not these two data-streams were used in simultaneous or step-wise approach would then only depend on computational time constraints. Note that using a two step approach can be a practical solution when one step requires longer simulations in order to optimize one or two “slow” process parameters. This study does not yet provide answers for how to deal with these issues globally, but it is an initial study on how to use biomass data in MDF while anticipating the availability of global datasets, e.g. from ESA BIOMASS mission.

The optimization helped to greatly decrease the uncertainty (reduction of 63% at Hesse and 78% at Le Bray) in total aboveground biomass projections in a future climate simulation at both sites (derived from the parameter uncertainties; see Supplementary Material, Section B). However, it only resulted in pronounced differences in the mean value at Le Bray (25 kg m^{-2} in 2100 after the optimization instead of 15 kg m^{-2} before). This was due to a larger NPP production after the optimization due to decreases in autotrophic respiration and increase in sapwood allocation. These results therefore demonstrate the importance of having a good NPP estimate in order to have reliable estimates of the AGB_tot and its temporal evolution. Use of NPP in the optimization is challenging however, as it cannot be measured directly and therefore is derived from other observations. There is thus a need for reliable estimates for autotrophic respiration, as has also been stated in other studies (Richardson et al., 2010), in order to derive the net carbon assimilated into the plant and soil.

Many of the large scale LSMs perform poorly when simulating aboveground biomass (Wolf et al., 2011). We have shown that using aboveground biomass increment data can help to constrain the allocation-related parameters, which therefore helped to improve the model-data fit and helped to reduce uncertainty in future projections. However, it is also possible that too simplistic allocation

scheme is contributing to LSM poor performance (De Kauwe et al., 2014). At large scales, the use of simplified schemes can be justified; however, the allocation between aboveground and belowground components is quite important as the decay processes for the corresponding biomass stores are very different (Repo et al., 2011). Some large scale models do not even separate between the aboveground and belowground wood biomass stores (Wolf et al., 2011), as this has not been essential in their developmental needs. We suggest that it is important to differentiate between these two stores and to try to constrain the allocation related parameters in order to achieve realistic simulations of biomass. Finally, our study also brings forth the importance of taking the forest age into account when studying biomass stocks into the future which many global LSMs do not do.

5. Conclusions

In this study we included aboveground forest biomass data in the optimization of the detailed and complex ORCHIDEE process-based model for the first time. Joint assimilation of annual aboveground increment data in addition to the micrometeorological flux data proved to be feasible (and beneficial) at two different forest sites, allowing the further constraint of allocation parameters in addition to parameters related to the “fast” processes in the model (photosynthesis, respiration and phenology). Adding total stand biomass revealed inconsistencies between these observations and the structure of this version of ORCHIDEE due to missing processes in the model. The “AR5” version of ORCHIDEE used in our study does not include a description of age-dependent growth rates, competition, and mortality processes that are crucial to simulate stand biomass. A new version of ORCHIDEE by Naudts et al. (2015) includes some of these missing processes, and therefore opens interesting avenues for the assimilation of aboveground biomass data in the future.

Aside from the model developments described above, our results highlight the need to have better differentiation between total ecosystem respiration and autotrophic respiration and to have a model that is consistent with different types of biomass observations. Having reliable observation-based NPP estimate would also be very useful for future MDF applications. The optimization we did in this work greatly helped to reduce uncertainties in the future climate scenarios for the aboveground biomass (shown in Supplement B).

Finally, we expect that having a greater number of observations on the parameters would help provide further information to the optimization by restricting their bounds. Clearly this is not a trivial task, as many direct observations of many parameters do not exist. However, we hope that closer communication between modellers and experimentalists taking site-based measurements would contribute to making this more feasible task.

Acknowledgements

This work is part of the EU-funded project Carbo-Extreme (FP7, GA 226701). TT would like to acknowledge the funding from the Finnish Academy (grant number 266803). We would like to acknowledge European Commission FP7 EMBRACE project, under Grant Agreement number 282672. We thank the two anonymous reviewers for their comments that greatly helped to improve this manuscript.

Appendix A. Equations controlling mostly fast processes (P1)

This Appendix shows the equations including the parameters that have been optimized in the MDF. The ORCHIDEE model equations are described in more detail in Krinner et al. (2005), Santaren et al. (2014) and Verbeeck et al. (2011). The optimized parameters are represented in bold.

Leaf area and phenology

The leaf area index (LAI, unit $m^2 m^{-2}$) is regulated by the PFT-specific maximum LAI (LAI_{max}) threshold after which carbon is not allocated to leaves anymore. The actual leaf biomass (unit $g C m^{-2}$) is obtained by multiplying LAI with the specific leaf area parameter (SLA) (Krinner et al., 2005). In the beginning of the growing season the deciduous forests use the carbohydrate reserve to grow leaves. The coniferous forests have no carbohydrate reserve in the ORCHIDEE model (Krinner et al., 2005). $K_{lai, happy}$ is the threshold that determines the LAI value after which the carbohydrate reserve will no longer be used, and it is multiplicative value of LAI_{max} . $\tau_{leafinit}$ is the number of days it takes for the plants to develop initial foliage using the carbohydrate reserve. The beginning of the leaf growth is dependent on the temperature threshold based on a growing degree day sum, and $K_{pheno, crit}$ is the optimized scalar value that modifies this threshold. $L_{age, crit}$ is the average critical age for the leaves and influences the photochemical efficiency of the foliage. Also the loss of leaf biomass is a function of the leaf age, with older leaves losing more biomass (Krinner et al., 2005). Leaf senescence is affected by air temperature and $c_{T, senescence}$, which is an offset term in the second order function describing the temperature dependency of the leaf senescence.

Photosynthesis

The photosynthesis for C3 plants is modelled in ORCHIDEE according to Farquhar et al. (1980). The ratio between the maximum carboxylation efficiency ($V_{c(max)}$) and the potential maximum rate of electron transport is assumed to be constant, the latter being two times $V_{c(max)}$. The actual $V_{c(max)}$ is dependent on leaf age, air temperature and soil water limitation (Verbeeck et al., 2011). The relative leaf age, which is the fraction of the critical leaf age ($L_{age, crit}$), controls the leaf efficiency. The temperature dependency is a function of optimal (T_{opt}), maximum (T_{max}) and minimum temperature (T_{min}) that are defined as:

$$T_{opt} = c_{T, opt} + b_{T, opt} \cdot T_l + a_{T, opt} \cdot T_l^2 \quad (A1)$$

$$T_{max} = c_{T, max} + b_{T, max} \cdot T_l + a_{T, max} \cdot T_l^2 \quad (A2)$$

$$T_{min} = c_{T, min} + b_{T, min} \cdot T_l + a_{T, min} \cdot T_l^2 \quad (A3)$$

where a_{T_i} , b_{T_i} and c_{T_i} are parameters and T_l is the long-term governing temperature. The soil water dependency of $V_{c(max)}$ is a function that includes the parameter f_{stress} , that sets the soil moisture limit under which $V_{c(max)}$ will be decreased. Hum_{cste} is the parameter that describes the exponential root profile that is used in the calculation for the water fraction available to the plant.

The stomatal conductance (g_s) is calculated with the Ball et al. (1987) model:

$$g_s = \frac{G_{S, slope} \cdot A \cdot h_r}{C_a} + g_{offset} \quad (A4)$$

where $G_{S, slope}$ is the slope of the stomatal conductance versus the net carbon assimilation rate A , h_r is the relative humidity (%), C_a is the atmospheric CO_2 concentration and g_{offset} is an offsetting parameter.

Autotrophic respiration

The maintenance respiration (R_m) is a function of air temperature (T) and biomass (B_i) of each separate biomass department i (Ruimy et al., 1996):

$$R_m^i = c(T_i) \cdot B_i \quad (A5)$$

$$c(T_i) = \max \left(c_0^i \cdot (MR_b + MR_a \cdot T), 0 \right) \quad (A6)$$

MR_a and MR_b are the coefficients of the linear temperature relationship and c_0 is a parameter specific for each biomass pool i . Growth respiration (R_g) is a constant fraction (GR_{frac}) of the assimilates available for growth:

$$R_g = GR_{frac} \cdot (A - R_m) \quad (A7)$$

Heterotrophic respiration

The representation of heterotrophic respiration in ORCHIDEE is based on a CENTURY type model (Parton et al., 1988). Soil litter is divided into structural, metabolic and woody litter pools that all have below- and aboveground compartments. The soil organic matter consists of three different pools: active, passive and slow. Each of these nine pools is decomposed according to:

$$R_h = K_{soilC} \sum_s \frac{\alpha_s}{\tau_s} f_{H, s} f_{T, s} B_s \quad (A8)$$

where R_h is the heterotrophic respiration, K_{soilC} is a multiplicative factor for the magnitude of the carbon pools, α_s is a pool specific coefficient partitioning heterotrophic respiration into pools, $f_{H, s}$ and $f_{T, s}$ are inhibition factors for slowing down decomposition due to soil moisture or low temperatures and B_s is the size of each soil carbon pool.

The temperature dependency of heterotrophic respiration is:

$$f_{T, surf/litter} = \min \left(1, Q_{10}^{\frac{T_{surf/litter} - 30}{10}} \right) \quad (A9)$$

where Q_{10} is an optimized parameter and temperature $T_{surf/litter}$ is the surface temperature for aboveground litter compartments and litter temperature for belowground litter compartments.

The exponential profile of litter decomposers is parameterized by a scaling depth Z_{decomp} that controls the effect of soil temperature on the decomposers. The effective temperature of litter decomposition T_{litter} is thus:

$$T_{litter} = \frac{\int_0^{Z_{litter}} T_{soil}(z) e^{-\frac{z}{Z_{decomp}}} dz}{1 - e^{-\frac{Z_{litter}}{Z_{decomp}}}} \quad (A10)$$

where Z_{litter} is the litter depth and z is depth of the vertical profile.

The soil/litter humidity inhibition factor f_H is:

$$f_{H, soil, litter} = \max \left(HR_{H, min}, \min \left(1, HR_{H, a} \cdot H_{soil, litter}^2 + HR_{H, b} \cdot H_{soil, litter} + HR_{H, c} \right) \right) \quad (A11)$$

where $H_{soil, litter}$ is the effective decomposition humidity of the aboveground soil/litter pools and $HR_{H, min}$, $HR_{H, a}$, $HR_{H, b}$ and $HR_{H, c}$ are fitted parameters.

The scaling depth Z_{decomp} has also influence on the soil/litter humidity (Verbeeck et al., 2011; Santaren et al., 2014). Litter humidity H_l is integrated through vertical depth to provide the H_{litter} according to:

$$H_{litter} = \frac{\int_0^{Z_{litter}} H_l(z) e^{-\frac{z}{Z_{decomp}}} dz}{1 - e^{-\frac{Z_{litter}}{Z_{decomp}}}} \quad (A12)$$

and H_l is exponential profile of the height of the dry reservoir h_{dry} of the double-bucket hydrological scheme (Choisnel, 1977), described as

$$H_l(z) = e^{-\frac{h_{dry}}{H_{crit_litter}}} \quad (A13)$$

where H_{crit_litter} is a scaling depth.

Energy balance

Radiation in ORCHIDEE can be reflected by vegetation, soil, dead leaves on ground or snow. The energy balance is expressed as:

$$R_{LW} + (1 - K_{albedo,veg} \cdot albedo_{veg} - albedo_{soil,snow,deadleaves}) \cdot R_{SW} - \varepsilon \sigma T_s^4 = ET + H + G \quad (A14)$$

where R_{LW} , R_{SW} , ε and σ are longwave and shortwave incoming radiations, the emissivity and the Stefan-Boltzmann constant, respectively. ET , H and G are evapotranspiration, sensitive and ground heat fluxes. The albedo of vegetation is multiplied by $K_{albedo,veg}$.

Aerodynamic resistance

The resistance of transfer for matter and energy between the vegetation and the measurement plane is described with the aerodynamic resistance r_a as:

$$r_a = \frac{1}{V_{wind} \cdot C_d(Z_{overheight})} \quad (A15)$$

where V_{wind} is the windspeed norm and C_d is the surface drag coefficient that is dependent on characteristic rugosity length $Z_{overheight}$ (Ducoudré et al., 1993).

Appendix B. Equations controlling mostly slow allocation processes (P2)

The carbon allocation scheme of ORCHIDEE is based on work by Friedlingstein et al. (1999) with some modifications by Krinner et al. (2005). It is a function of moisture, light and modelled nitrogen availability. These different availabilities all range between zero and one, with zero meaning no limitation. The light availability A_l is a function of LAI, in a way that increasing LAI increases light limitation. Belowground limitation A_b is the minimum of the moisture and nitrogen limitation. The belowground moisture limitation is calculated from relative soil moisture of different soil layers and supposing an exponential decrease in root profile with depth. The nitrogen limitation is calculated using moisture and temperature conditions.

The light availability and belowground limitation are used to calculate the preliminary allocation fractions for leaves, fine roots and sapwood ($f_{pr,l}$, $f_{pr,r}$ and $f_{pr,s}$) as

$$f_{pr,r} = \max\left(r_{min}, r_0 \frac{3A_l}{A_l + 2A_b}\right), \quad (B1)$$

$$f_{pr,s} = s_0 \frac{3A_b}{2A_l + A_b}, \quad (B2)$$

$$f_{pr,l} = \max(a_{min}, \min(a_{max}, 1 - f_{pr,s} - f_{pr,r})), \quad (B3)$$

where $r_{min} = 0.15$, $a_{min} = 0.2$, $a_{max} = 0.5$ and the prior value for both s_0 and r_0 is 0.3, but these two parameters are optimized. After this the preliminary root fraction is calculated again, using:

$$f_{pr,r} = 1 - f_{pr,s} - f_{pr,l}. \quad (B4)$$

To calculate the final allocation fractions, the preliminary allocation fractions are used with some modifications for special cases.

If the LAI value exceeds the PFT-specific maximum LAI (LAI_{max}), no carbon is allocated to leaves and this carbon is allocated to the sapwood instead. Deciduous trees will allocate the whole assimilate to the carbohydrate reserve during their senescence period. During the growing season 10% of the assimilate will be allocated to the fruit biomass pool (Sitch et al., 2003), even though this ratio is likely to be a large overestimation (Bellassen et al., 2010).

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.agrformet.2016.12.004>.

References

- Ahlström, A., Schurgers, G., Arneth, A., Smith, B., 2012. Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environ. Res. Lett.* 7, 044008.
- Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A.S., Martin, P.H., Berbigier, P., Bernhofer, Ch., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., Vesala, T., 2000. Estimates of the annual net carbon and water exchange of European forests: the EUROFLUX methodology. *Adv. Ecol. Res.* 30, 113–175.
- Aubinet, M., Berbigier, P., Bernhofer, C., Cescatti, A., Feigenwinter, C., Granier, A., Grünwald, Th., Havrankova, K., Heinesch, B., Longdoz, B., Marcolla, B., Montagnani, L., Sedlak, P., 2005. Comparing CO₂ storage and advection conditions at night at different Euroflux sites. *Bound.-Lay. Meteorol.* 116, 63–93.
- Avitabile, V., Herold, M., Heuvelink, G.B.M., Lewis, S.L., Phillips, O.L., Asner, G.P., Armston, J., Ashton, P.S., Banin, L., Bayol, N., Berry, N.J., Boeckx, P., de Jong, B.H.J., DeVries, B., Girardin, C.A.J., Kearsley, E., Lindsell, J.A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E.T.A., Nagy, L., Oje, L., Quinones, M.J., Ryan, C.M., Ferry, S.J.W., Sunderland, T., Laurin, G.V., Gatti, R.C., Valentini, R., Verbeeck, H., Wijaya, A., Willcock, S., 2016. An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biol.* 22, 1406–1420. <http://dx.doi.org/10.1111/gcb.13139>.
- Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., Houghton, R.A., 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon density maps. *Nat. Climate Change* 2, 182–185. <http://dx.doi.org/10.1038/nclimate1354>.
- Bacour, C., Peylin, P., MacBean, N., Rayer, P., Delage, F., Santaren, D., 2015. Joint assimilation of eddy-covariance flux measurements and satellite observations within a process-oriented biosphere model. *J. Geophys. Res. Biogeosci.* 120, 1839–1857. <http://dx.doi.org/10.1002/2015JG002966>.
- Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. In: Biggins, J. (Ed.), *A Model Predicting Stomatal Conductance and Its Contribution to the Control of Photosynthesis Under Different Environmental Conditions*, vol. 4. M. Nijhoff Publishers, Dordrecht, pp. 221–224.
- Bellassen, V., Le Maire, G., Dhôte, J.F., Ciais, P., Viovy, N., 2010. Modelling forest management within a global vegetation model – Part 1: Model structure and general behavior. *Ecol. Model.* 221, 2458–2474.
- Bellassen, V., Le Maire, G., Guin, O., Dhôte, J.F., Ciais, P., Viovy, N., 2011. Modelling forest management within a global vegetation model – part 2: model validation from a tree to a continental scale. *Ecol. Model.* 222, 57–75.
- Berbigier, P., Bonnefond, J.M., Mellmann, P., 2001. CO₂ and water vapour fluxes for 2 years above Euroflux forest site. *Agric. For. Meteorol.* 108, 183–197.
- Bolin, B., Rodhe, H., 1973. A note on the concepts of age distribution and transit time in natural reservoirs. *Tellus B* 25, 58–62.
- Bréda, N., Huc, R., Granier, A., Dreyer, E., 2006. Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. *Ann. For. Sci.* 63, 625–644.
- Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* 11, 335–355. <http://dx.doi.org/10.1111/j.1365-2486.2005.00897.x>.
- Carvalhais, N., Reichstein, M., Seixas, J., Collatz, G.J., Pereira, J.S., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M.J., Valentini, R., 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. *Global Biogeochem. Cycl.* 22, <http://dx.doi.org/10.1029/2007GB003033>, GB2007.
- Carvalhais, N., Reichstein, M., Ciais, P., Collatz, G.J., Mahecha, M.D., Montagnani, L., Papale, D., Rambal, S., Seixas, J., 2010. Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints. *Global Change Biol.* 16, 2813–2829. <http://dx.doi.org/10.1111/j.1365-2486.2010.02173.x>.
- Carvalhais, N., Forkel, M., Khomik, M., Bellarby, J., Jung, M., Migliavacca, M., Mu, M., Saatchi, S., Santoro, M., Thurner, M., Weber, U., Ahrens, B., Beer, C., Cescatti, A., Randerson, J.T., Reichstein, M., 2014. Global covariation of carbon turnover

- times with climate in terrestrial ecosystems. *Nature* 514, 213–217, <http://dx.doi.org/10.1038/nature13731>.
- Choisnel, E., 1977. Le bilan d'énergie et le bilan hydrique du sol. *Météorologie* 6, 103–133.
- Ciais, P., Schelhaas, M.J., Zaehle, S., Piao, S.L., Cescatti, A., Liski, J., Luysaert, S., LeMaire, G., Schulze, E.D., Bouriaud, O., Freibauer, A., Valentini, R., Nabuurs, G.J., 2008. Carbon accumulation in European forests. *Nat. Geosci.* 1, 425–429.
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R.B., Piao, S., Thornton, P., 2013. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Carbon and Other Biogeochemical Cycles. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Curtis, P.S., Hanson, P.J., Bolstad, P., Barford, C., Randolph, J.C., Schmid, H.P., Wilson, K.B., 2002. Biometric and eddy-covariance based estimates of annual carbon balance storage in five eastern North American deciduous forests. *Agric. For. Meteorol.* 113, 3–19.
- De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Walker, A.P., Dietze, M.C., Wang, Y.-P., Luo, Y., Jain, A.K., El-Masri, B., Hickler, T., Wärlind, D., Weng, E., Parton, W.J., Thornton, P.E., Wang, S., Prentice, I.C., Asao, S., Smith, B., McCarthy, H.R., Iversen, C.M., Hanson, P.J., Warren, J.M., Oren, R., Norby, R.J., 2014. Where does the carbon go? A model–data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO₂ enrichment sites. *New Phytol.* 203, 883–899, <http://dx.doi.org/10.1111/nph.12847>.
- Ducoudré, N.L., Laval, K., Perrier, A., 1993. SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model. *J. Climate* 6, 248–273.
- Epron, D., Farque, L., Lucot, E., Badot, P.M., 1999. Soil CO₂ efflux in a beech forest: dependence on soil temperature and soil water content. *Ann. For. Sci.* 56, 221–226.
- Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthesis CO₂ fixation in leaves of C₃ species. *Planta* 149, 78–90.
- Friedlingstein, P., Joel, G., Field, C.B., Fung, I.Y., 1999. Toward an allocation scheme for global terrestrial carbon models. *Global Change Biol.* 5, 755–770, <http://dx.doi.org/10.1046/j.1365-2486.1999.00269.x>.
- Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., Knutti, R., 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate* 27, 511–526, <http://dx.doi.org/10.1175/JCLI-D-12-00579.1>.
- Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Publishing Company (412 pp).
- Granier, A., Biron, P., Lemoine, D., 2000. Water balance, transpiration and canopy conductance in two beech stands. *Agric. For. Meteorol.* 100, 291–308.
- Granier, A., Bréda, N., Longdoz, B., Gross, P., Ngao, J., 2008. Ten years of fluxes and stand growth in a young beech forest at Hesse, North-eastern France. *Ann. For. Sci.* 64, 704.
- Grote, R., Kiese, R., Grünwald, T., Ourcival, J.-M., Granier, A., 2011. Modelling forest carbon balances considering tree mortality and removal. *Agric. For. Meteorol.* 151, 179–190.
- Haupt, R.L., Haupt, S.E., 2004. *Practical Genetic Algorithms*, 2nd edition. John Wiley & Sons, Hoboken, New Jersey The United States (253 pp).
- Hollinger, D., Richardson, A., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Phys.* 25, 873–885.
- Hurt, G.C., Fisk, J., Thomas, R.Q., Dubayah, R., Moorcroft, P.R., Shugart, H.H., 2010. Linking models and data on vegetation structure. *J. Geophys. Res.* 115, G00E10, <http://dx.doi.org/10.1029/2009JG000937>.
- Jarosz, N., Brunet, Y., Lamaud, E., Irvine, M., Bonnefond, J.-M., Loustau, D., 2008. Carbon dioxide and energy flux partitioning between the understorey and the overstorey of a maritime pine forest during a year with reduced soil water availability. *Agric. For. Meteorol.* 148, 1508–1523.
- Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., Mathieu, P.-P., 2012. Consistent assimilation of MERIS FAPAR and atmospheric CO₂ into a terrestrial vegetation model and interactive mission benefit analysis. *Biogeosciences* 9, 3173–3184, <http://dx.doi.org/10.5194/bg-9-3173-2012>.
- Keenan, T.F., Davidson, E., Moffat, A.M., Munger, W., Richardson, A.D., 2012. Using model-data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial ecosystem carbon cycling. *Global Change Biol.* 18, 2555–2569, <http://dx.doi.org/10.1111/j.1365-2486.2012.02684.x>.
- Keenan, T.F., Davidson, E., Munger, W., Richardson, A.D., 2013. Rate my data: quantifying the value of ecological data for the development of models of the terrestrial carbon cycle. *Ecol. Appl.* 23, 273–286.
- Knorr, W., Kaminski, T., Scholze, M., Gobron, N., Pinty, B., Giering, R., Mathieu, P.-P., 2010. Carbon cycle data assimilation with a generic phenology model. *J. Geophys. Res.* 115, <http://dx.doi.org/10.1029/2009JG001119>, G04017.
- Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C., 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycle* 19, <http://dx.doi.org/10.1029/2003GB002199> (GB1015).
- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., Richardson, A.D., 2012. Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences* 9, 3757–3776.
- Lasslop, G., Reichstein, M., Kattge, J., Papale, D., 2008. Influences of observation errors in eddy flux data on inverse model parameter estimation. *Biogeosciences* 5, 1311–1324, <http://dx.doi.org/10.5194/bg-5-1311-2008>.
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P., Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biol.* 16, 187–208, <http://dx.doi.org/10.1111/j.1365-2486.2009.02041.x>.
- MacBean, N., Maignan, F., Peylin, P., Bacour, C., Ciais, P., Bréon, F.-M., 2015. Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences* 12, 7185–7208.
- MacBean, N., Peylin, P., Chevallier, F., Scholze, M., Schürmann, G., 2016. Consistent assimilation of multiple data streams in a carbon cycle data assimilation system. *Geosci. Model Dev.* 9, 3569–3588, <http://dx.doi.org/10.5194/gmd-9-3569-2016>.
- Medlyn, M., Loustau, D., Delzon, S., 2002. Temperature response of parameters of a biochemically based model of photosynthesis I. Seasonal changes in mature maritime pine (*Pinus pinaster* Ait.). *Plant Cell Environ.* 25, 1165–1166.
- Montpiéd, P., Granier, A., Dreyer, E., 2009. Seasonal time-course of gradients of photosynthetic capacity and mesophyll conductance to CO₂ across a beech (*Fagus sylvatica* L.) canopy. *J. Exp. Bot.* 60, 2407–2418.
- Naudts, K., Ryder, J., McGrath, M.J., Otto, J., Chen, Y., Valade, A., Bellasien, V., Berhongaray, G., Bönsch, G., Campioli, M., Ghattas, J., De Groot, T., Haverd, V., Kattge, J., MacBean, N., Maignan, F., Merilä, P., Penuelas, J., Peylin, P., Pinty, B., Pretzsch, H., Schulze, E.D., Solyga, D., Vuichard, N., Yan, Y., Luysaert, S., 2015. A vertically discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the energy, water and carbon fluxes. *Geosci. Model Dev.* 8, 2035–2065.
- Ogée, J., Brunet, Y., Loustau, D., Berbigier, P., Delzon, S., 2003. MuSICA, a CO₂, water and energy multilayer, multileaf pine forest model: evaluation from hourly to yearly time scales and sensitivity analysis. *Global Change Biol.* 9, 697–717.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., Yakir, D., 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences* 3, 571–583, <http://dx.doi.org/10.5194/bg-3-571-2006>.
- Parton, W., Stewart, J., Cole, C., 1988. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry* 5, 109–131, <http://dx.doi.org/10.1007/BF02180320>.
- Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P.J., Kuppel, S., Koffi, E.N., Kane, A., Maignan, F., Chevallier, F., Ciais, P., Prunet, P., 2016. A new step-wise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle. *Geosci. Model Dev.* 9, 3321–3346, <http://dx.doi.org/10.5194/gmd-9-3321>.
- Pietsch, S.A., Hasenauer, H., 2006. Evaluating the self-initialization procedure for large-scale ecosystem models. *Global Change Biol.* 12, 1658–1669.
- Porté, A., Trichet, P., Bert, D., Loustau, D., 2002. Allometric relationships for branch and tree woody biomass of Maritime pine (*Pinus pinaster* Aét.). *For. Ecol. Manag.* 158, 71–83.
- Purves, D., Pacala, S., 2008. Predictive models of forest dynamics. *Science* 320, 1452–1453.
- Raj, R., Hamm, N.A.S., Tol, C.V.D., Stein, A., 2016. Uncertainty analysis of gross primary production partitioned from net ecosystem exchange measurements. *Biogeosciences* 13, 1409–1422, <http://dx.doi.org/10.5194/bg-13-1409-2016>.
- Rayner, P.J., 2010. The current state of carbon-cycle data assimilation. *Curr. Opin. Environ. Sustain.* 2, 289–296.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, A., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.* 11, 1424–1439.
- Repo, A., Tuomi, M., Liski, J., 2011. Indirect carbon dioxide emissions from producing bioenergy from forest harvest residues. *GCB Bioenergy* 3, 107–115.
- Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J.P., Dail, D.B., Davidson, E.A., Scott, N.A., Evans, R.S., Hughes, H., Lee, J.T., Rodrigues, C., Savage, K., 2010. Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. *Oecologia* 164, 25–40.
- Richardson, A.D., Aubinet, M., Barr, A.G., Hollinger, D.Y., Ibrom, A., Lasslop, G., Reichstein, M., 2012. Uncertainty quantification. In: *Aubinet, M., Vesala, T., Papale, D. (Eds.), Eddy Covariance – A Practical Guide to Measurement and Data Analysis*. Springer (438 pp).
- Ruimy, A., Kergoat, L., Field, C.B., Saugier, B., 1996. The use of CO₂ flux measurements in models of the global terrestrial carbon budget. *Global Change Biol.* 2, 287–296, <http://dx.doi.org/10.1111/j.1365-2486.1996.tb00080.x>.
- Saatchi, S., Harris, N.L., Brown, S., Lefsky, M., Mitchard, E.T., Salas, W., Zutta, B.R., Buermann, W., Lewis, S.L., Hagen, S., Petrova, S., White, L., Silman, M., Morel, A., 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9899–9904.
- Santaren, D., Peylin, P., Viovy, N., Ciais, P., 2007. Optimizing a process-based ecosystem model with eddy-covariance flux measurements: a pine forest in southern France. *Global Biogeochem. Cycle* 21, <http://dx.doi.org/10.1029/2006GB002834> (GB2013).
- Santaren, D., Peylin, P., Bacour, C., Ciais, P., Longdoz, B., 2014. Ecosystem model optimization using in-situ flux observations: benefit of monte-carlo vs variational schemes and analyses of the year-to-year model performances. *Biogeosciences* 11, 7137–7158.

- Santoro, M., Beaudoin, A., Beer, C., Cartus, O., Fransson, J.E., Hall, R.J., Schmullius, C., Schepaschenko, D., Shvidenko, A., Thurner, M., Wegmueller, U., 2015. **Forest growing stock volume of the northern hemisphere: spatially explicit estimates for 2010 derived from Envisat ASAR. Remote Sens. Environ.** 168, 316–334.
- Schürmann, G.J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., Zaehle, S., 2016. Constraining a land surface model with multiple observations by application of the MPI-carbon cycle data assimilation system. *Geosci. Model Dev.* 9, 2999–3026. <http://dx.doi.org/10.5194/gmd-9.2999-2016>.
- Schwalm, C.R., Williams, C.A., Schaefer, K., Anderson, R., Arain, M.A., Baker, I., Barr, A., Black, T.A., Chen, G., Chen, J.M., Ciais, P., Davis, K.J., Desai, A., Dietze, M., Dragoni, D., Fischer, M.L., Flanagan, L.B., Grant, R., Gu, L., Hollinger, D., Izaurralde, R.C., Kucharik, C., Lafleur, P., Law, B.E., Li, L., Li, Z., Liu, S., Lokupitiya, E., Luo, Y., Ma, S., Margolis, H., Matamala, R., McCaughey, H., Monson, R.K., Oechel, W.C., Peng, C., Poulter, B., Price, D.T., Riciutto, D.M., Riley, W., Sahoo, A.K., Sprintsin, M., Sun, J., Tian, H., Tonitto, C., Verbeeck, H., Verma, S.B., 2010. A model-data intercomparison of CO₂ exchange across north america: results from the north american carbon program site synthesis. *J. Geophys. Res.* 115, G00H05. <http://dx.doi.org/10.1029/2009JG001229>.
- Sitch, S., Smith, B., Prentice, I.C., Arneeth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S., 2003. **Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model.** *Global Change Biol.* 9, 161–185.
- Tarantola, A., 2005. **Inverse Problem Theory and Methods for Model Parameter Estimation.** SIAM, The U.S (342 pp).
- Thornton, P.E., Law, B., Gholz, H., Clark, K., Falge, E., Ellsworth, D., Goldstein, A.H., Monson, R.K., Hollinger, D., Falk, M., Chen, J., Sparks, J.P., 2002. **Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests.** *Agric. For. Meteorol.* 113, 185–222.
- Thurner, M., Beer, C., Santoro, M., Carvalhais, N., Wutzler, T., Schepaschenko, D., Shvidenko, A., Kompter, E., Ahrens, B., Levick, S.R., Schmullius, C., 2014. Carbon stock and density of northern boreal and temperate forests. *Global Ecol. Biogeogr.* 23, 297–310. <http://dx.doi.org/10.1111/geb.12125>.
- Verbeeck, H., Samson, R., Granier, A., Montpied, P., Lemeur, R., 2008. **Multi-year analysis of GPP in a temperate beech forest in France.** *Ecol. Model.* 210, 85–103.
- Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., Ciais, P., 2011. Seasonal patterns of CO₂ fluxes in Amazon forests: fusion of eddy covariance data and the ORCHIDEE model. *J. Geophys. Res.* 116, <http://dx.doi.org/10.1029/2010JG001544> (G02018).
- Wang, Y.P., Baldocchi, D., Leuning, R., Falge, E., Vesala, T., 2007. Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight FLUXNET sites. *Global Change Biol.* 13, 652–670. <http://dx.doi.org/10.1111/j.1365-2486.2006.01225.x>.
- Wang, T., Breder, P., Ciais, P., Piao, S., Mahecha, M.D., Chevallier, F., Reichstein, M., Ottlé, C., Maignan, F., Arain, A., Bohrer, G., Cescatti, A., Kiely, G., Law, B.W., Lutz, M., Montagnani, L., Moors, E., Osborne, B., Panferov, O., Papale, D., Vaccari, F.P., 2012. **State-dependent errors in a land surface model across biomes inferred from eddy covariance observations on multiple timescales.** *Ecol. Model.* 246, 11–25.
- Williams, M., Richardson, A.D., Reichstein, M., Stoy, P.C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D.Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C.M., Wang, Y.–P., 2009. Improving land surface models with FLUXNET data. *Biogeosciences* 6, 1341–1359. <http://dx.doi.org/10.5194/bg-6-1341-2009>.
- Wolf, A., Ciais, P., Bellassen, V., Delbart, N., Field, C.B., Berry, J.A., 2011. Forest biomass allometry in global land surface models. *Global Biogeochem. Cycles* 25, <http://dx.doi.org/10.1029/2010GB003917> (GB3015).
- Wu, S.H., Jansson, P.-E., Kolari, P., 2011. **Modeling seasonal course of carbon fluxes and evapotranspiration in response to low temperature and moisture in a boreal Scots pine ecosystem.** *Ecol. Model.* 222, 3103–3119.
- Wutzler, T., Carvalhais, N., 2014. Balancing multiple constraints in model-data integration: weights and the parameter block approach. *J. Geophys. Res. Biogeosci.* 119, 2112–2129. <http://dx.doi.org/10.1002/2014JG002650>.
- Wutzler, T., Wirth, C., Schumacher, J., 2008. **Generic biomass functions for Common beech (*Fagus sylvatica* L.) in Central Europe – predictions and components of uncertainty.** *Can. J. For. Res.* 48, 1661–1675.
- Zaehle, S., Friend, A.D., 2010. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochem. Cycle* 24, <http://dx.doi.org/10.1029/2009GB003521> (GB1005).
- Zaehle, S., Medlyn, B.E., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hickler, T., Luo, Y., Wang, Y.-P., El-Masri, B., Thornton, P., Jain, A., Wang, S., Warland, D., Weng, E., Parton, W., Iversen, C.M., Gallet-Budynek, A., McCarthy, H., Finzi, A., Hanson, P.J., Prentice, I.C., Oren, R., Norby, R.J., 2014. Evaluation of 11 terrestrial carbon-nitrogen cycle models against observations from two temperate free-air CO₂ enrichment studies. *New Phytol.* 202, 803–822. <http://dx.doi.org/10.1111/nph.12697>.