



# Bayesian inversions of a dynamic vegetation model at four European grassland sites

J. Minet<sup>1</sup>, E. Laloy<sup>2</sup>, B. Tychon<sup>1</sup>, and L. François<sup>3</sup>

<sup>1</sup>Université de Liège, Arlon Campus Environnement, Avenue de Longwy 185, 6700 Arlon, Belgium

<sup>2</sup>Belgian Nuclear Research Centre (SCK-CEN), Boerentang 200, 2400 Mol, Belgium

<sup>3</sup>Université de Liège, UMCCB, Allée du six août 17, 4000 Liège, Belgium

Correspondence to: J. Minet (julien.minet@ulg.ac.be)

Received: 15 January 2015 – Published in Biogeosciences Discuss.: 29 January 2015

Revised: 9 April 2015 – Accepted: 16 April 2015 – Published: 13 May 2015

**Abstract.** Eddy covariance data from four European grassland sites are used to probabilistically invert the CARAIB (CARbon Assimilation In the Biosphere) dynamic vegetation model (DVM) with 10 unknown parameters, using the DREAM<sub>(ZS)</sub> (DiffeREntial Evolution Adaptive Metropolis) Markov chain Monte Carlo (MCMC) sampler. We focus on comparing model inversions, considering both homoscedastic and heteroscedastic eddy covariance residual errors, with variances either fixed a priori or jointly inferred together with the model parameters. Agreements between measured and simulated data during calibration are comparable with previous studies, with root mean square errors (RMSEs) of simulated daily gross primary productivity (GPP), ecosystem respiration (RECO) and evapotranspiration (ET) ranging from 1.73 to 2.19, 1.04 to 1.56 g C m<sup>-2</sup> day<sup>-1</sup> and 0.50 to 1.28 mm day<sup>-1</sup>, respectively. For the calibration period, using a homoscedastic eddy covariance residual error model resulted in a better agreement between measured and modelled data than using a heteroscedastic residual error model. However, a model validation experiment showed that CARAIB models calibrated considering heteroscedastic residual errors perform better. Posterior parameter distributions derived from using a heteroscedastic model of the residuals thus appear to be more robust. This is the case even though the classical linear heteroscedastic error model assumed herein did not fully remove heteroscedasticity of the GPP residuals. Despite the fact that the calibrated model is generally capable of fitting the data within measurement errors, systematic bias in the model simulations are observed. These are likely due to model inadequacies such as shortcomings in the photosynthesis modelling. Besides the residual error treatment,

differences between model parameter posterior distributions among the four grassland sites are also investigated. It is shown that the marginal distributions of the specific leaf area and characteristic mortality time parameters can be explained by site-specific ecophysiological characteristics.

## 1 Introduction

Covering about 38 % of the European agricultural area and 8 % of the land surface (FAO, 2011), grassland is an important land cover class in Europe, which shows a wide range of different ecological characteristics. By stocking carbon, temperate grassland might play an important role in climate change mitigation in Europe (Soussana et al., 2004) and on the world scale (O'Mara, 2012). Large uncertainties, however, remain in the estimation of the (source or sink) carbon fluxes since those largely depend on farming management options.

In environmental modelling, grassland growth models have received less attention than the long-standing and highly developed crop models. Since grasslands are agroecosystems that can be considered either as agricultural or semi-natural lands, grassland models were designed for two main purposes: the simulation of forage and dairy or meat production, and the simulation of the carbon fluxes at the land–atmosphere interface. Several crop models were adapted for grassland growth modelling (e.g., STICS; Ruget, 2009; Dumont et al., 2014, EPIC; Williams et al., 2008), especially when the management of the grassland remained similar to crop management, i.e., when the grassland was used for tem-

porary forage production and was cut rather than grazed by animals. Some other models were specifically developed for grasslands (e.g., SPACSYS; Wu et al., 2007), sometimes coupled with animal production models (e.g., PaSim; Graux et al., 2013), whereas grassland models were also developed from dynamic vegetation models (DVMs) such as LPJmL (Bondeau et al., 2007), adapted from the LPJ model (Sitch et al., 2003). Being process-based models, DVMs are well suited for large-scale spatial simulations and can account for a wide range of current and projected climatic conditions.

To be used for simulation-based decision making, a DVM must be properly parametrized. Model parameter values can be derived from (1) laboratory experiments as, e.g., the stomatal conductance described by the Ball–Berry model (Ball et al., 1987), (2) in situ field measurements, or (3) model inversion using calibration data measurements or (4) spatialized databases (e.g., from remote sensing). Model inversion (also referred to as calibration) consists of automatically finding those model parameters that allow the model to adequately reproduce the available observed data. The collection of representative and high-quality data is thus of paramount importance for inversion, as DVMs require an adequate parametrization that is sufficiently representative of the range of conditions over the spatial extent of the simulation. Typically, DVMs use different sets of parameters that are assigned to specific vegetation classes that grow together over the same area or in geographically distinct biomes. Dynamic vegetation model inversion needs a sufficient number of sites with varying ecophysiological conditions that are supposed to be representative of the considered vegetation classes or biomes, but still well-delimited (Knorr and Katge, 2005). Model inversion using continuous, gridded data (e.g., from remote sensing; Patenaude et al., 2008) could also help in determining optimal parameters for large areas, but computation time can be a limiting factor for such application.

Given the high number of eddy covariance experimental sites across the world, eddy covariance measurements are particularly appealing for inversion of the DVMs (Friend et al., 2007). Furthermore, the long-standing rise in computational resources not only increased modelling capabilities in terms of temporal and spatial resolution but also opened new avenues for quantifying the uncertainty associated with the estimated model parameters and its effect on model simulations. In particular, the Bayesian framework for inverse modelling is increasingly used in the DVM community (e.g., Hartig et al., 2012). Bayesian methods such as Markov chain Monte Carlo (MCMC) sampling aim to derive a representative set of all parameter combinations that are consistent with the observed data and available prior information. This set of parameters is referred to as the posterior distribution.

Nevertheless, eddy covariance data are known to be associated with relatively large measurement uncertainties, implying both systematic and random errors (see Aubinet et al. (2012), chapter 7, for a comprehensive description of all sources of eddy covariance uncertainties). As eddy covari-

ance data are the result of a long process chain, they can be affected by instrumental measurement error (e.g., calibration and design errors), sampling errors due to the variability of the fluxes in time and space and data treatment error (e.g., due to the gap filling of missing data). Uncertainties in eddy covariance data are also strongly dependent on the time resolution of the fluxes, tending to diminish with time aggregation (Richardson and Hollinger, 2005). It is crucial to account for these random data uncertainties in the inversion since an improper statistical treatment can cause the parameter posterior distribution to be strongly biased (e.g., Fox et al., 2009). Quantifying random eddy covariance data errors is not straightforward (Hollinger and Richardson, 2005; Lasslop et al., 2008), but these errors are typically characterized by a variance that is proportional to the magnitude of the data, i.e., they show heteroscedasticity (e.g., Lasslop et al., 2008). Therefore, it has been suggested (Richardson et al., 2008) that the measurement error variance can be modelled as a linear function of the magnitude of the flux with a non-null intercept, as random errors are non-null even when the flux equals 0. However, while the random error can be taken into account in the inversion, systematic measurement errors can only be removed by instrument calibration.

In this study, data from eddy covariance stations over four grassland sites are inverted for the CARAIB (CARbon Assimilation In the Biosphere) dynamic vegetation model parameters within a Bayesian framework. This is both the first automatic calibration of the CARAIB model and its first application to managed grassland modelling, which required the adaptation of the model to grass cutting and grazing. The main objective is to compare the modelling of the carbon and water fluxes over the four grassland sites using four different ways of treating the eddy covariance data errors during the inversion. Both homoscedastic and heteroscedastic residual error models are considered, either fixed beforehand or sampled along with the model parameters. A second objective is then to compare the site-specific posterior parameter distributions obtained for the four grasslands, given their climatic, ecological and management characteristics.

## 2 Materials and methods

### 2.1 Experimental sites and data

In this study, we focus on four long-term experimental sites (see Table 1) that are semi-natural permanent grasslands: Grillenburg, Germany (Prescher et al., 2010); Oensingen (intensively managed), Switzerland (Ammann et al., 2007); Monte Bondone, Italy, (Wohlfahrt et al., 2008) and Laqueuille (extensively managed), France, (Klumpp et al., 2011). The four sites pertain to the global FLUXNET network and, as such, a large number of studies were conducted using eddy covariance data from these sites. The FLUXNET website (<http://fluxnet.ornl.gov/>) provides lists of references per site.

The four sites are located in western and central Europe and experience different climate, altitude, soil and management conditions. They can be classified according to the De Martonne–Gottman aridity index, which is inversely related with the site aridity. Oensingen is the most intensively managed site and the only one that is fertilized (about  $200 \text{ kg N ha}^{-1} \text{ yr}^{-1}$ ). The other three sites are extensively managed, with no organic or mineral fertilization. The last two sites are mid-mountainous grassland, while the first two sites are situated at a lower altitude. Only the grassland in Laqueuille is grazed by animals during the growing season, while the other three are hay meadows that are cut once or several times a year. Note that, although grass cutting occurred on the 13 June 2005 in Grillenburg according to the given management data, it was not observed in the measured eddy covariance fluxes because of gap filling of missing data. As a result, this cut was neglected in the modelling.

The four grasslands are equipped with eddy covariance stations for measuring ecosystem fluxes. Flux measurements and field data sets were made available through a coordinated task of the FACCE/MACSUR (Food Agriculture Climate Change/Modeling European Agriculture with Climate Change for food Security) knowledge hub, which aims at performing an intercomparison of grassland models (Ma et al., 2014) by running several grassland models with the same field data sets collected under various climatic and management conditions. Field data sets hold the necessary information for feeding the grassland model: hourly meteorological records of climatic variables, soil physical parameters, management information such as cutting dates or grazing charges, and initial conditions. Daily eddy covariance data included net ecosystem exchange (NEE;  $\text{g C m}^{-2} \text{ day}^{-1}$ ), gross primary productivity (GPP;  $\text{g C m}^{-2} \text{ day}^{-1}$ ), ecosystem respiration (RECO;  $\text{g C m}^{-2} \text{ day}^{-1}$ ) and evapotranspiration (ET;  $\text{mm day}^{-1}$ ). It is worth noting that only the NEE and ET are directly measured by the eddy covariance station (i.e., fluxes of  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , respectively) and that GPP and RECO are derived from these measurements.

In this study, only GPP, RECO and ET measurements were used in the inverse modelling. Adding NEE measurements would be ineffective as they are directly dependent on GPP and RECO. GPP and RECO were used since they are directly linked with the photosynthesis and respiration processes, respectively, while the influence of these two processes is mixed in the NEE measurements. Other combinations including the NEE were first tested but resulted in poorer agreement between measured and modelled data. The full data range including gap-filled data was inverted, since these data are gap-filled according to specific protocols that are standards in the eddy covariance community.

## 2.2 The CARAIB model

### 2.2.1 Description of the model

CARAIB is a physically based dynamic vegetation model that was developed for the simulation of the carbon cycle on the global scale (Warnant et al., 1994; Nemry et al., 1996; Otto et al., 2002). It calculates the carbon fluxes through the soil–vegetation–atmosphere continuum by simulating eco-physiological processes: photosynthesis, carbon allocation to plant pools, and autotrophic and heterotrophic respiration. The CARAIB model has been used in numerous paleoclimatology, vegetation and crop modelling studies. The reader is referred to the aforementioned references for a full model description.

For  $\text{C}_3$  plants, photosynthesis is computed according to the model of Farquhar et al. (1980). The stomatal conductance governing the flux of  $\text{CO}_2$  through the stomata is described on the leaf scale with the Ball–Berry approach (Ball et al., 1987), using the model of Leuning (1995) with further adaptations from Van Wijk et al. (2000) to account for soil water stress affecting the stomatal conductance. Photosynthesis and respiration processes are computed at 2-hour time steps on a half-day basis, and the model assumes a symmetry with respect to solar noon time; that is, computation of these processes is made for half the day and further aggregated using a daily time step. Other processes, e.g., related to soil hydrology or carbon allocation, are computed on a daily basis.

In this study, a single plant functional type (PFT) is considered (BAG 22 as defined in Laurent et al., 2004, 2008) corresponding to the flora that can be encountered in European grasslands, i.e., species of Poaceae and Asteraceae. The model was adapted for simulating the grassland sites by adding management functions for grass cutting and grazing. Grass cutting is modelled by the removal of a part of the plant carbon mass so that the model matches given values of leaf area index after cutting. Grazing is modelled such that a given fraction of the plant carbon mass is removed every day according to the grazing charge. The dates of the grass cutting and the duration of the grazing periods were known and fixed in the simulations. Daily meteorological data recorded at the experimental sites were used in the model, i.e., minimal and maximal temperature, precipitation, solar radiation, relative air humidity, and wind velocity. Although they can affect vegetation modelling (Gottschalk et al., 2007; Rivington et al., 2006; Zhao et al., 2012), uncertainties in the meteorological data were not considered in this study.

Thirty-three parameters per PFT are set in CARAIB. These parameters govern photosynthesis, plant physiology process (e.g., specific leaf area, carbon-to-nitrogen ratio), allocation of carbon and residence times in the different pools of carbon, including plants and soil pools, land surface–atmosphere interactions (albedo, roughness length), and tolerance to extreme conditions (thresholds and response times). During the model development, parameter values in

CARAIB were mainly taken from the literature (Warnant, 1999) and further compared with observed values (remote sensing, field data and paleorecords). So far, no model inversions were performed with the CARAIB model.

### 2.2.2 Choice of parameters

In this study, 10 model parameters were sampled (Table 2). They were chosen according to their presupposed importance – that is, the model sensitivity to these parameters – and because some parameter values were already known in the measured data from the experimental sites. Default values that were defined during the model development and used in previous research are given in Table 2. These parameters govern the main processes of the model, namely, the photosynthesis, the respiration and carbon transfer between carbon pools:

- The slope  $g_1$  and the intercept  $g_0$  ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) of the stomatal conductance as described in Leuning (1995) are directly related to the photosynthesis since they govern the stomatal conductance. They are thus related to the gross primary productivity (GPP) and evapotranspiration (ET) with respect to the meteorological conditions. While most of ecological models, including CARAIB, use an empirical approach for stomatal conductance, derived from the Ball–Berry model, Medlyn et al. (2011) recently reconciled the empirical approach with the theoretical background based on the optimal stomatal behaviour (Farquhar et al., 1980), which states that there is a trade-off for stomata between maximizing carbon gain (photosynthesis) and minimizing water loss (transpiration). These new developments in the theoretical understanding of the empirical relationship push forward the necessity to measure or calibrate the stomatal conductance parameters under different environmental conditions. Although single values of these parameters are used for regional or global modelling of  $C_3$  plant photosynthesis (e.g., Sitch et al., 2008), it is known that stomatal conductance parameters actually vary through time and space according to the environmental conditions and plant species.
- The specific leaf area (SLA;  $\text{m}^2 \text{g C}^{-1}$ ) is defined in CARAIB as the leaf area per unit of carbon mass of the plants. It is used in the model to convert the assimilated mass of carbon into leaf area index. Besides its role in the model, SLA is often studied as a plant trait that is used for predicting the plant resource use strategy or for clustering plants species into functional groups. Maximizing the photosynthesis while minimizing leaf respiration, high-SLA leaves (thin leaves) are productive but also more vulnerable and short-lived (Wilson et al., 1999). They are thus better adapted to resource-rich environment, where leaves can be quickly reconstructed (Poorter and De Jong, 1999). On the other hand, low-SLA leaves (thick leaves) are often encountered

in drought-adapted (Marcelis et al., 1998) or shade-tolerant species (Evans and Poorter, 2001) and for the lower, self-shaded leaves of a plant. SLA is also known to vary over the course of the season and according to the leaf age (Wilson et al., 1999). Nevertheless, the concept of SLA is sometimes problematic for some plant species with complex plant geometry (Vile et al., 2005), e.g., highly folded leaves or with a non-negligible part of the photosynthetic tissues located on the stem, as is the case among the Poaceae species. In these simulations, SLA is defined for the PFT that is supposed to represent European grasslands, and, therefore, SLA should actually be considered as an effective parameter among the grassland species and for the whole plant body.

- The characteristic mortality time (year) of the plant in normal  $\tau$  and in stress conditions  $\tau_s$  is, respectively, the characteristic time for the renewal of the plant ( $\tau$ ) and the time it takes for the plant to die in stress conditions ( $\tau_s$ ). The stress conditions occur when temperatures reach either low or high extreme values, for soil water content below a certain threshold or for low irradiance values. The default values were 0.667 year for  $\tau$ , meaning a renewal of the plant within 8 months, and 0.083 year for  $\tau_s$ , meaning a characteristic mortality time in stress conditions of 1 month.
- Two carbon-to-nitrogen ratios are defined for the photosynthetic active carbon pool of the plant ( $C/N_1$ ) and for the remainder of the plant ( $C/N_2$ ). The nitrogen content of the leaves play a crucial role in the photosynthesis, and increasing nitrogen content (decreasing  $C/N$ ) fosters photosynthetic activity. A low  $C/N$  ratio in plant usually occurs together with high nitrogen content in soils, that is, a resource-rich environment.
- Three parameters govern the rates of the soil heterotrophic respiration:  $\gamma_1$  for the respiration of the “green litter“,  $\gamma_2$  for the respiration of the “non-green litter” and  $\gamma_3$  for the respiration of the soil organic carbon.

## 2.3 Probabilistic inversion methodology

### 2.3.1 Inverse problem

To acknowledge that measurements and modelling errors are inevitable, the inverse problem is commonly represented by the stochastic relationship

$$F(\mathbf{z}) = \mathbf{d} + \mathbf{e}, \quad (1)$$

where  $F$  is a deterministic, error-free forward model that expresses the relation between the uncertain parameters  $\mathbf{z}$  and the measurement data  $\mathbf{d}$  and where the noise term  $\mathbf{e}$  lumps measurement and model errors.

**Table 1.** Grassland sites and periods of simulations.

	Coordinates	Altitude	Management	Fertili- sation	De Martonne– Gottman index	Calibration years	Validation years
Grillenburg, DE	13.50° E, 50.95° N	380 m	cutting (1–3 yr <sup>-1</sup> )	no	32	2004–2006	2007–2008
Oensingen, CH	7.73° E, 47.28° N	450 m	cutting (3–5 yr <sup>-1</sup> )	yes	38	2002–2005	2006–2008
Monte Bondone, IT	11.03° E, 46.00° N	1500 m	cutting (1 yr <sup>-1</sup> )	no	35	2003–2005	2006–2007
Laqueuille, FR	2.73° E, 45.63° N	1040 m	grazing	no	41	2004–2007	2008–2010

Inversions were performed within a Bayesian framework, which treats the unknown model parameters  $\mathbf{z}$  as random variables with the posterior probability density function (pdf)  $p(\mathbf{z}|\mathbf{d})$  given by

$$p(\mathbf{z}|\mathbf{d}) = \frac{p(\mathbf{z})p(\mathbf{d}|\mathbf{z})}{p(\mathbf{d})} \propto p(\mathbf{z})L(\mathbf{z}|\mathbf{d}), \quad (2)$$

where  $p(\mathbf{z})$  denotes the prior distribution of  $\mathbf{z}$  and  $L(\mathbf{z}|\mathbf{d}) \equiv p(\mathbf{d}|\mathbf{z})$  signifies the likelihood function of  $\mathbf{z}$ . The normalization factor  $p(\mathbf{d}) = \int p(\mathbf{z})p(\mathbf{d}|\mathbf{z})d\mathbf{z}$  is obtained from numerical integration over the parameter space so that  $p(\mathbf{z}|\mathbf{d})$  scales to unity. The quantity  $p(\mathbf{d})$  is generally difficult to estimate in practice but is not required for parameter inference. In the remainder of this study, we will focus on the unnormalized posterior  $p(\mathbf{z}|\mathbf{d}) \propto p(\mathbf{z})L(\mathbf{z}|\mathbf{d})$ . For numerical stability, it is often preferable to work with the log-likelihood function,  $\ell(\mathbf{z}|\mathbf{d})$ , instead of  $L(\mathbf{z}|\mathbf{d})$ . If we assume the error  $\mathbf{e}$  to be normally distributed, uncorrelated and with an unknown constant variance,  $\sigma^2$ , the log-likelihood function can be written as

$$\ell(\mathbf{z}|\mathbf{d}) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^N [d_i - F_i(\mathbf{z})]^2, \quad (3)$$

where  $\sigma$  can be fixed beforehand or sampled jointly with the other model parameters  $\mathbf{z}$ .

The homoscedasticity (i.e., constant variance) assumption for  $\mathbf{e}$  may be excessively strong in many cases. Considering the residual errors,  $\mathbf{e}$ , to be heteroscedastic, Eq. (3) becomes

$$\ell(\mathbf{z}|\mathbf{d}) = -\frac{N}{2}\log(2\pi) - \sum_{i=1}^N \log(\sigma_i) - \frac{1}{2}\sum_{i=1}^N \frac{[d_i - F_i(\mathbf{z})]^2}{\sigma_i^2}, \quad (4)$$

where the  $\sigma_i$  represents the individual residual error standard deviations that can be gathered into a vector  $\boldsymbol{\sigma}$ . Here also,  $\boldsymbol{\sigma}$  can either be fixed beforehand or sampled along with  $\mathbf{z}$  (see Sect. 2.3.4).

### 2.3.2 Multi-objective likelihood function

In this work, we chose three types of eddy covariance data for the calibration:  $\mathbf{d}_1$  (GPP),  $\mathbf{d}_2$  (RECO) and  $\mathbf{d}_3$  (ET). We further assume that the corresponding residual errors,  $\mathbf{e}_1$ ,  $\mathbf{e}_2$  and  $\mathbf{e}_3$ , are uncorrelated, leading to the following multi-objective log-likelihood function:

$$\ell(\mathbf{z}|\mathbf{d}_{1,2,3}) = \ell(\mathbf{z}|\mathbf{d}_1) + \ell(\mathbf{z}|\mathbf{d}_2) + \ell(\mathbf{z}|\mathbf{d}_3). \quad (5)$$

The weighting between the three components of  $\ell(\mathbf{z}|\mathbf{d}_{1,2,3})$  is an important issue. The constant ( $\sigma$ ) and non-constant ( $\sigma_i$ ) standard deviations in Eqs. (3) and (4), respectively, basically weight the respective influences of  $\mathbf{e}_1$ ,  $\mathbf{e}_2$  and  $\mathbf{e}_3$  on the log likelihood defined by Eq. (5). Distinct homoscedastic or heteroscedastic residual error models must be specified for  $\mathbf{e}_1$ ,  $\mathbf{e}_2$  and  $\mathbf{e}_3$ . This was done for both the homoscedastic and heteroscedastic cases either by specifying the residual error standard deviations beforehand or by jointly inferring these standard deviations along with the model parameters.

### 2.3.3 Homoscedastic and heteroscedastic error models

Based on prior knowledge of the measurement errors, the homoscedasticity assumption simply reduces to assigning values to  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  in Eqs. (3) and (5). These values were fixed to 3 g C m<sup>-2</sup> day<sup>-1</sup> for the GPP measurements, 1.5 g C m<sup>-2</sup> day<sup>-1</sup> for the RECO measurements and 1 mm day<sup>-1</sup> for the ET measurements. As stated earlier, measurement errors associated with eddy covariance fluxes are, however, typically found to be heteroscedastic, with a variance that is assumed to be linearly related to the magnitude of the measured data (Richardson et al., 2008):

$$\sigma_{d,i} = \frac{1}{2}\sigma_{0,d} \left( \frac{d_i}{\bar{d}} + 1 \right), \quad (6)$$

where the variable  $d$  denotes either GPP, RECO, or ET measurements,  $i = 1, \dots, N$  are measurement times and  $\sigma_{0,d}$  is equivalent to  $\sigma_1$ ,  $\sigma_2$  or  $\sigma_3$  in the homoscedastic case. We refer to the inversions based on these homoscedastic and heteroscedastic error models as HO1 and HE1, respectively. It is worth noting that by fixing the standard deviations to known measurement errors, one implicitly assumes that the model is able to describe the observed system up to the observation errors. This might not be realistic in environmental modelling, where models are always fairly simplified descriptions of a much more complex reality.

### 2.3.4 Joint inference of the homoscedastic and heteroscedastic error model parameters

Still under the Gaussianity assumption, a more advanced treatment of the residual error models considers the simultaneous inference of the standard deviations with the model parameters, i.e., it considers the standard deviation of the

residual errors as unknowns. Doing so assumes that residual errors are expected to be a mixture of both model (equations and inputs) and observational errors. For the homoscedastic case, this simply consists of jointly sampling  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  along with the model parameters,  $z$ .

The heteroscedastic error model then becomes

$$\sigma_{d,i} = a d_i + b, \quad (7)$$

where the  $a$  and  $b$  coefficients are to be jointly inferred with  $z$  from the measurement data. Using Eq. (7) thus leads to the addition of six variables to the sampling problem:  $a_1$ ,  $a_2$ ,  $a_3$ ,  $b_1$ ,  $b_2$  and  $b_3$ . We refer to the joint inversions of these homoscedastic and heteroscedastic error models as HO2 and HE2, respectively. In these inversions, a total predictive uncertainty around the model values can be computed by adding to the modelled data a random noise drawn from a normal distribution with mean 0 and standard deviation  $\sigma$  sampled from its posterior distribution (HO2) or computed by Eq. (7; HE2).

The simultaneous inference of model parameters with homoscedastic or heteroscedastic error model parameters requires the definition of their prior probability distributions. Based on the available prior information, uniform (flat) priors are used for the 10 model parameters contained in  $z$  (see Table 2). We follow two guidelines for specifying the prior densities of the error model parameters. First, we would like to obtain posterior standard deviations that are as small as possible within the range permitted by the model and measurement data errors in order to get the lowest possible data misfits. Second, the magnitudes of the different prior distributions should reflect the desired weights of the different data types within the multi-objective inference. These weights translate the modeller's relative preferences among the three modelling objectives in Eq. (5). We therefore use normal distributions with mean 0 truncated at 0 to avoid negative values. The prescribed weights then correspond to the different standard deviations of these normal distributions:

$$p(X) = \frac{1}{\sigma_X B} \phi\left(\frac{X - \mu_X}{\sigma_X}\right) \propto \phi\left(\frac{X - \mu_X}{\sigma_X}\right), \quad (8)$$

where the  $X$  variable is either  $\sigma_j$ ,  $a_j$  or  $b_j$  for  $j = 1, 2, 3$ ; where the value of  $\sigma_X$  expresses the modeller's preference for objective  $j$  compared to the other objectives (the smaller  $\sigma_X$ , the larger the relative weight of objective  $j$ ); where  $\phi(\cdot)$  signifies the probability density function of the standard normal distribution; where  $\mu_X$  is set to 0 for maximizing the prior density of  $X$  towards small values; and where the constant  $B$  depends on the lower ( $v$ ) and upper ( $w$ ) limits of the truncation interval

$$B = \Phi\left(\frac{w - \mu_X}{\sigma_X}\right) - \Phi\left(\frac{v - \mu_X}{\sigma_X}\right), \quad (9)$$

in which  $\Phi(\cdot)$  denotes the cumulative distribution function of the standard normal distribution.

This treatment of multi-objective Bayesian inference is in line with the work of Reichert and Schuwirth (2012), who further considered different statistical models for model and observation errors. Overall, this resulted in four different ways of treating the eddy covariance data uncertainties: fixed homoscedastic (HO1) and heteroscedastic (HE1) error models and jointly inferred homoscedastic (HO2) and heteroscedastic (HE2) error models. Using the HO1 and HE1 models led to a total of 10 inferred parameters, whereas using the HO2 and HE2 models resulted into a total of 13 and 16 inferred parameters, respectively. Table 2 lists the marginal prior distributions used for all sampled parameters. The upper and lower bounds of these distributions were either set to their maximal and minimal physically possible values or determined on the basis of expert knowledge.

### 2.3.5 Markov chain Monte Carlo sampling

The goal of the inference is to estimate the posterior distribution  $p(z|\mathbf{d})$  where the 10-, 13- or 16-dimensional  $z$  vector contains all sampled parameters and  $\mathbf{d}$  signifies the conditioning data:  $\mathbf{d} = \{d_{1,2,3}\}$  herein. As an exact analytical solution of  $p(z|\mathbf{d})$  is not available, we resort to Markov chain Monte Carlo (MCMC) simulation to generate samples from this distribution. The basis of this technique is a Markov chain that generates a random walk through the search space and iteratively finds parameter sets with stable frequencies stemming from the posterior pdf of the model parameters (see, e.g., Robert and Casella, 2004, for a comprehensive overview of MCMC simulation).

The MCMC sampling efficiency strongly depends on the assumed proposal distribution used to generate transitions in the Markov chain. In this work, the state-of-the-art DREAM<sub>(ZS)</sub> (ter Braak and Vrugt, 2008; Vrugt et al., 2009; Laloy and Vrugt, 2012) (DiffeRential Evolution Adaptive Metropolis) algorithm is used to generate posterior samples. A detailed description of this sampling scheme including convergence proof can be found in the literature cited and is thus not reproduced herein.

Convergence of the MCMC sampling to the posterior distribution is monitored by means of the potential scale reduction factor of Gelman and Rubin (1992),  $\hat{R}$ . For each parameter of interest, this statistic compares the average within-chain variance to the variance of all the chains mixed together. The smaller the difference between these two variances, the closer to 1 the value of the  $\hat{R}$  diagnostic. Values of  $\hat{R}$  smaller than 1.2 are commonly deemed to indicate convergence to a stationary distribution. In this study, posterior distributions of the parameters were drawn from the point where all parameters achieved  $\hat{R} < 1.2$ . This is more conservative than the conventional practice of stopping the inference when  $\hat{R} < 1.2$  for every parameter. The mean acceptance rate of the proposed samples, AR (%), is an important sampling property and is thus also reported. An excessively small fraction of accepted candidate points indicates poor

**Table 2.** Default values and prior distributions of the 10 model parameters and prior distributions of the statistical parameters of the homoscedastic and heteroscedastic error models. The label U means a uniform distribution, TG signifies a zero-mean Gaussian distribution truncated at 0 to avoid negative values and SD denotes the prescribed standard deviation of a TG distribution.

Parameter	Units	Default value	Prior type	Range	SD
Model parameters					
g1		9	U	[1, 20]	N/A*
g0	mol m <sup>-2</sup> s <sup>-1</sup>	0.01	U	[0.005, 0.03]	N/A
SLA	m <sup>2</sup> g C <sup>-1</sup>	0.025	U	[0.01, 0.08]	N/A
τ	year	0.667	U	[0.5, 2]	N/A
τ <sub>s</sub>	year	0.0833	U	[0.01, 0.5]	N/A
C/N1		16	U	[5, 40]	N/A
C/N2		32	U	[10, 80]	N/A
γ1		20	U	[5, 40]	N/A
γ2		10	U	[5, 40]	N/A
γ3		0.2	U	[0, 1]	N/A
Homoscedastic error model parameters (for HO2 inversions only)					
σ <sub>GPP</sub>	g C m <sup>-2</sup> day <sup>-1</sup>	N/A	TG	[0, 54]	9
σ <sub>RECO</sub>	g C m <sup>-2</sup> day <sup>-1</sup>	N/A	TG	[0, 27]	4.5
σ <sub>ET</sub>	mm day <sup>-1</sup>	N/A	TG	[0, 18]	3
Heteroscedastic error model parameters (for HE2 inversions only)					
a <sub>GPP</sub>		N/A	TG	[0, 27 × $\overline{Y_{GPP}}$ ]	4.5 × $\overline{Y_{GPP}}$
a <sub>RECO</sub>		N/A	TG	[0, 13.5 × $\overline{Y_{RECO}}$ ]	2.25 × $\overline{Y_{RECO}}$
a <sub>ET</sub>		N/A	TG	[0, 9 × $\overline{Y_{ET}}$ ]	1.5 × $\overline{Y_{ET}}$
b <sub>GPP</sub>	g C m <sup>-2</sup> day <sup>-1</sup>	N/A	TG	[0, 27]	4.5
b <sub>RECO</sub>	g C m <sup>-2</sup> day <sup>-1</sup>	N/A	TG	[0, 13.5]	2.25
b <sub>ET</sub>	mm day <sup>-1</sup>	N/A	TG	[0, 9]	1.5

\* Not applicable.

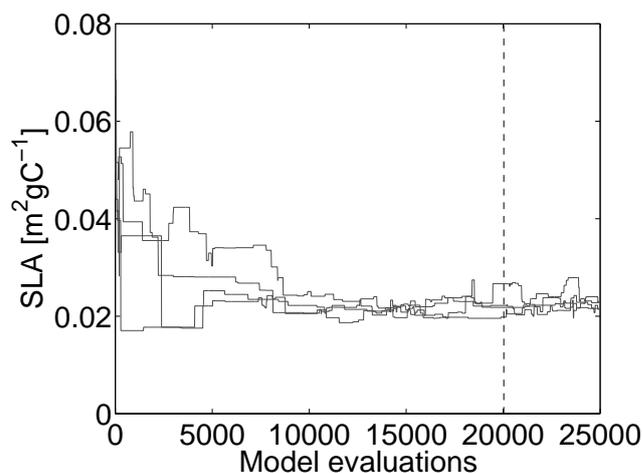
mixing of the chains due to too wide a proposal distribution. In contrast, a very large acceptance rate signals too narrow a proposal distribution, causing the chains to remain in the close vicinity of their current locations. The optimal value for AR depends on the proposal and target distributions, but a range of 10–30 % generally indicates good performance of DREAM<sub>(ZS)</sub>.

### 3 Results

#### 3.1 Parameter estimation

##### 3.1.1 Parameter samplings and convergence of the algorithm

The DREAM<sub>(ZS)</sub> algorithm was run with four parallel chains, initialized by sampling the prior parameter distribution (Table 2). As an example, Fig. 1 shows sampling trajectories of DREAM<sub>(ZS)</sub> parametrized with four chains for the SLA parameter and inversion HO1 at the Oensingen site. The  $\hat{R}$  convergence statistic becomes < 1.2 for each parameter after about 20 000 forward model runs, and the AR over the last



**Figure 1.** Sampled values of the specific leaf area (SLA) by DREAM<sub>(ZS)</sub> parametrized with four chains for the Oensingen site and the fixed homoscedastic error model (inversion HO1). The vertical dashed line indicates when convergence has been reached according to the  $\hat{R}$  statistic.

50 % model evaluations is about 18 %. Overall, convergence was achieved for all MCMC trials after some 15 000–30 000 forward runs with AR values in the range of 10–30 %, except for the inversions associated with the Laqueuille site that showed AR values as low as 5 %.

### 3.1.2 Posterior parameter distributions

Figure 2 presents marginal posterior histograms of the 10 model parameters for all experimental sites, considering the inferred homoscedastic error model (inversion HO2). In the remainder of this document, results are mainly detailed for this inversion scenario, since it generally led to the lowest data misfit statistics in calibration. For some parameters (e.g., SLA and C/N1), the marginal posterior distributions are narrow compared to the prior parameter range. This indicates a large sensitivity of the model to the considered parameter. In contrast, some other parameters such as  $\gamma_2$  are poorly resolved, demonstrating a relative insensitivity. Asymmetric edge-hitting distributions are also observed such as for C/N1 and C/N2 in Monte Bondone. In a Bayesian inversion of eddy covariance data obtained from a forest site, Braswell et al. (2005) found that 7 out of 26 marginal parameter distributions were edge-hitting. Extending the prior parameter ranges would lead to unphysical or implausible parameter values. Edge-hitting distributions reveal model inadequacies and/or large systematic measurements errors. For some parameters, posterior distributions were fairly distinct from the default values that were used in previous studies (Table 2), such as high  $g_1$  values. Values of the characteristic mortality time  $\tau$  also generally increased compared to the default value.

Table 3 shows the most likely parameter values for the four experimental sites; these parameter values resulted in the highest values of the log-likelihood function. Some of the parameters present contrasting values between inversion scenarios and/or experimental sites, which may be related to the different ecological characteristics of the sites as discussed in section 4.3. Depending on the width of the posterior distributions, the most likely parameter values are well resolved or largely uncertain. As a result, a comparison between the experimental sites must account for the posterior distributions of the parameters.

## 3.2 Measured and modelled carbon and water fluxes with calibration data

### 3.2.1 Measured and modelled data in Monte Bondone

As the parameter sampling resulted in posterior distributions of the parameters instead of single values, ensembles of posterior modelled signals can be represented as a graph. In Fig. 3, measured and modelled eddy covariance data are depicted for the experimental site of Monte Bondone for inversions with the inferred homoscedastic error model (inversion

HO2). The posterior ranges of the modelled signals are represented by the dark grey shaded areas for the prediction uncertainty due to parameter uncertainties and by the light grey shaded areas for the total predictive uncertainty (at 95 % confidence level). This total prediction uncertainty is computed using the standard deviation of the residual errors  $\sigma$  as sampled by the inversions and, therefore, cannot be computed for the NEE. The site of Monte Bondone was chosen here since there is one single cut a year (indicated by the vertical arrows in Fig. 3) that is clearly identifiable, which facilitates the interpretation of the fluxes. The dates of cutting corresponded to a sudden drop in the GPP in the middle of the year, which was followed by a gradual increase. They were also observed in the NEE graphs, with a sudden increase in the NEE.

There was overall good agreement between measured and modelled signals. It is worth noting that the posterior ranges of modelled data were not constant over time and were not related to the magnitude of the signals. The ranges due to parameter uncertainties were relatively small and did not encompass the measured data. Overall, it could be observed that measured eddy covariance data have stronger dynamics than the modelled signals, meaning that the CARAIB model cannot follow the fast fluctuations of the GPP (and other signals) over time. In particular, the model could not simulate the highest peaks in GPP well.

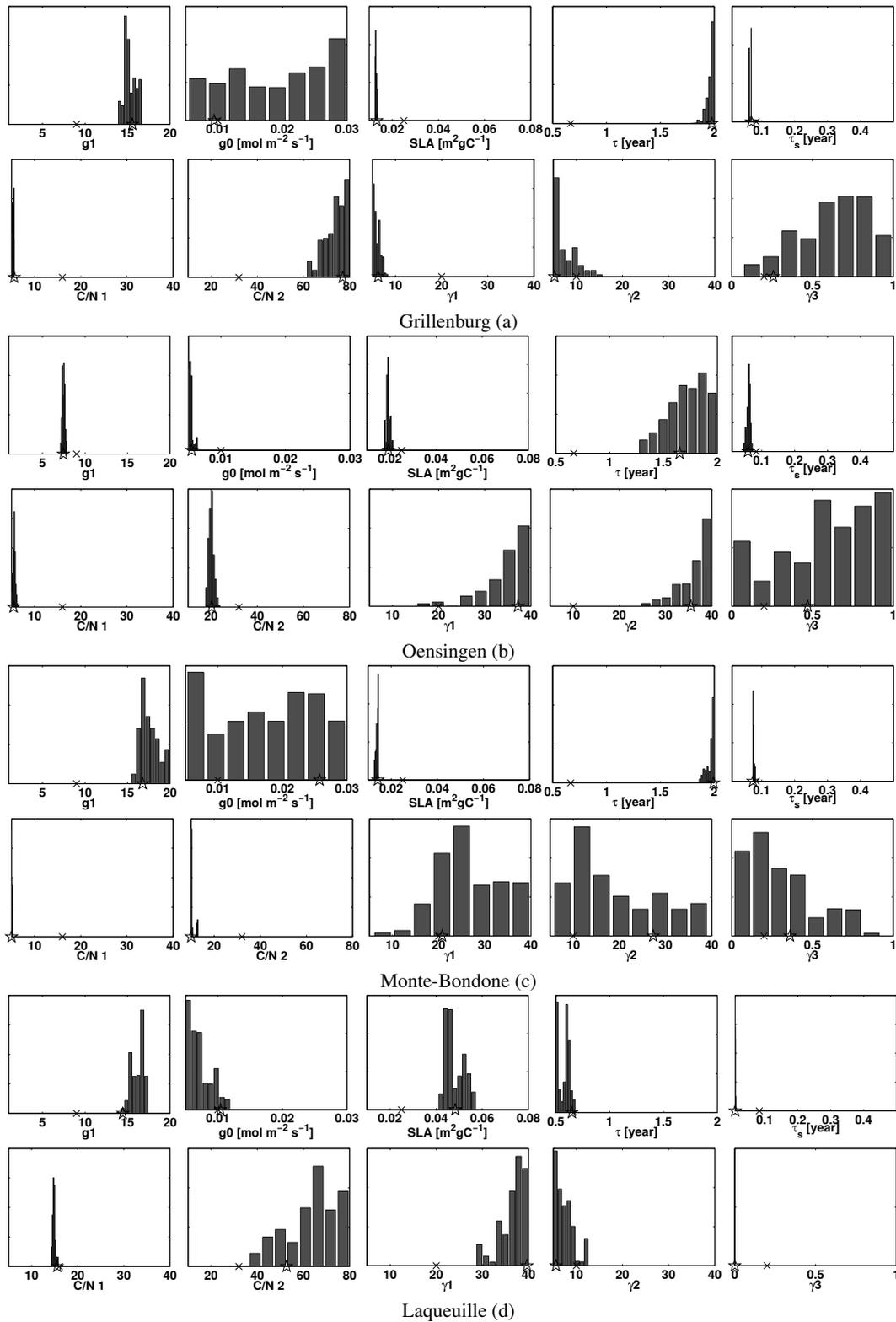
### 3.2.2 Measured and modelled data across sites

Considering the other three experimental sites (Fig. 4), there was a similar agreement between measured and modelled signals, although the sites displayed different behaviour in terms of GPP as their management varies: there are several cuts per year in Grillenburg and Oensingen, while Laqueuille is a grazed meadow. In general, the peaks in GPP cannot be simulated well by the model. The modelled GPP seemed averaged out when compared to the measured signals, as observed before in Monte Bondone (Fig. 3a).

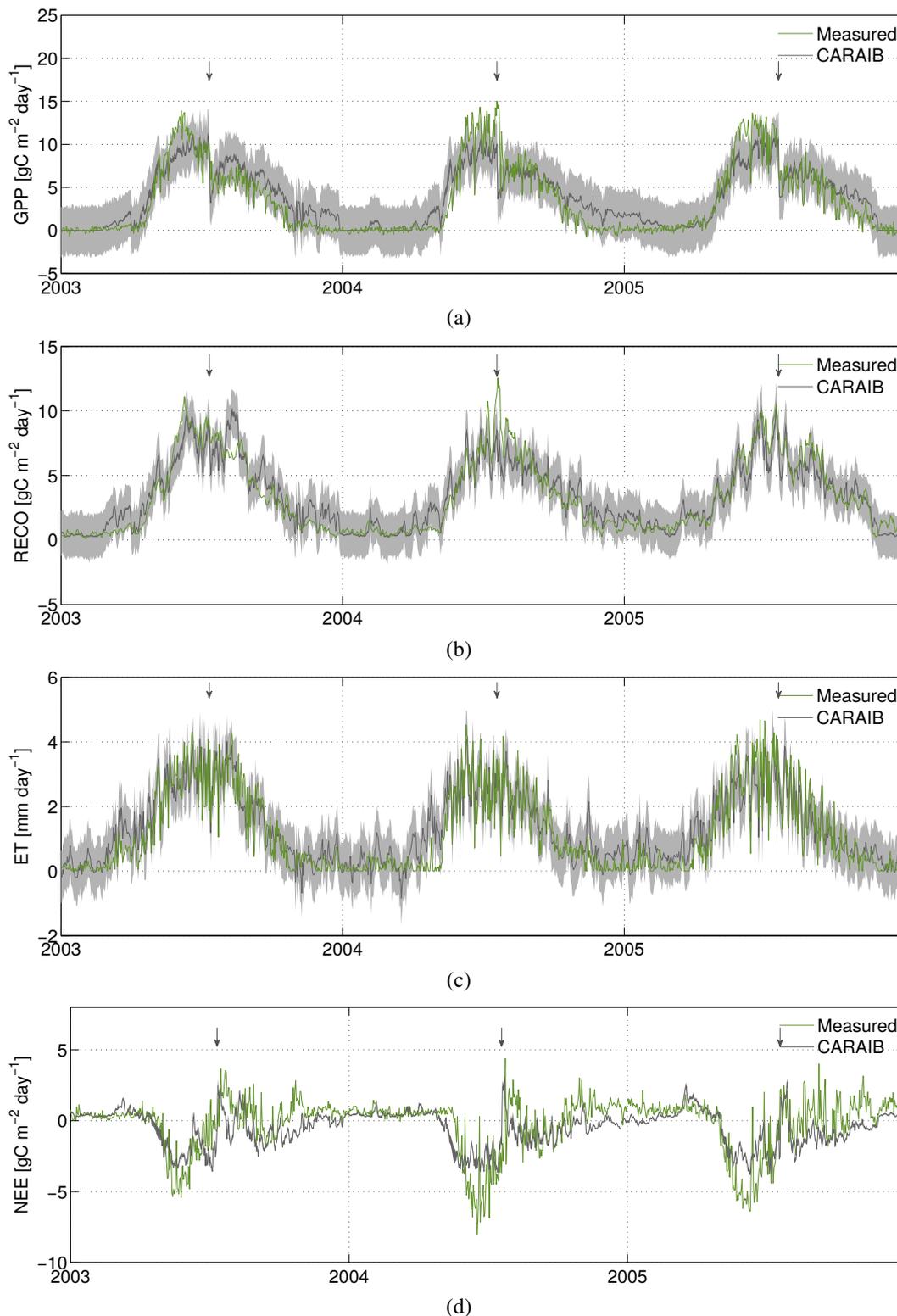
All the graphical comparisons between measured and modelled signals could not be shown but are summarized in Table 4 for the homoscedastic and heteroscedastic cases, and with the fixed and inferred error models, using the root mean square error (RMSE), the  $R^2$  and the Nash and Sutcliffe (1970) model efficiency criterion ( $E$ ) between measured and modelled signals. The latter criterion takes values from  $-\infty$  to 1. A value of 1 means a perfect match between measurements and model simulations, a value of 0 indicates that the mean of the observed data is as accurate as the modelled values, and an efficiency less than 0 occurs when the mean of the observed data reproduces the observations better than the modelled values. The maximum log-likelihood value “ml” that was obtained by the algorithm is also indicated. Note that performance criteria were also computed for the NEE, although these data were not used in the model inversions. Overall, the best agreement was found for the Monte Bondone site and the worst for the Laqueuille site. The low-

**Table 3.** Most likely CARAIB model parameter values for all inversion scenarios.

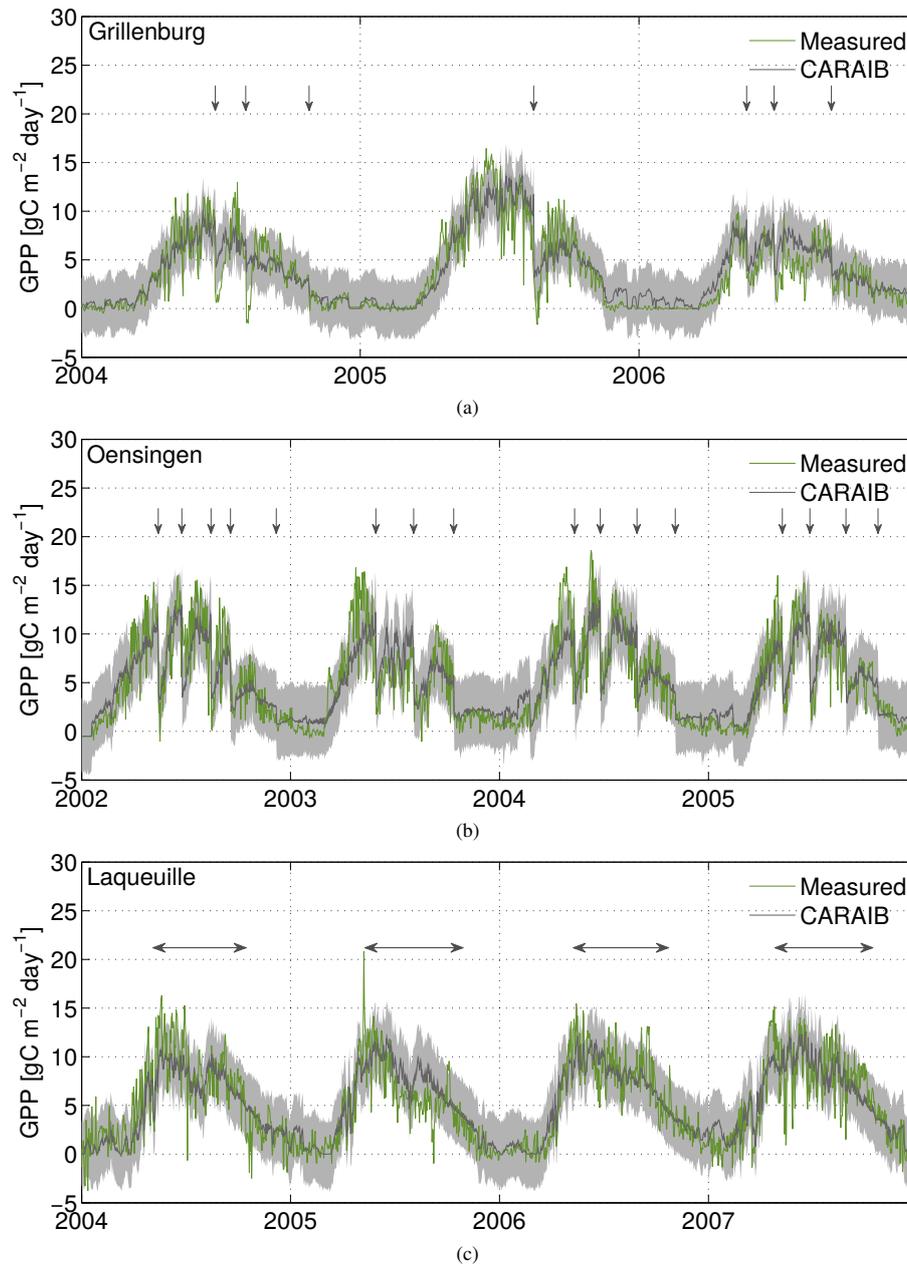
	Grillenburg	Oensingen	Monte Bondone	Laqueuille
Fixed homoscedastic error model inversions (HO1)				
g1	16.8	7.3	18.8	18.6
g0 (mol m <sup>-2</sup> s <sup>-1</sup> )	0.0265	0.00507	0.00637	0.0248
SLA (m <sup>2</sup> g C <sup>-1</sup> )	0.0126	0.0234	0.0155	0.0197
τ (year)	1.99	1.27	1.98	1.49
τ <sub>s</sub> (year)	0.0861	0.0526	0.0212	0.023
C / N1	5	6.69	5.02	5.43
C / N2	78.6	19.9	10.6	11
γ1	5.07	39.1	38.2	26.1
γ2	5.1	39.9	38.8	36.9
γ3	0.73	0.507	0.421	1.49 × 10 <sup>-5</sup>
Fixed heteroscedastic error model inversions (HE1)				
g1	3.45	8	19.8	20
g0 (mol m <sup>-2</sup> s <sup>-1</sup> )	0.027	0.00544	0.0297	0.0299
SLA (m <sup>2</sup> g C <sup>-1</sup> )	0.0161	0.0151	0.0142	0.0191
τ (year)	1.96	1.7	1.96	0.746
τ <sub>s</sub> (year)	0.0202	0.0687	0.0153	0.0234
C / N1	5.11	5.1	5	5
C / N2	77.9	20.3	10.2	10
γ1	8.09	39.5	31.4	38.8
γ2	5.96	37.4	30.9	24.8
γ3	0.358	0.806	0.981	0.688
Inferred homoscedastic error model inversions (HO2)				
g1	15.6	7.46	16.8	14.5
g0 (mol m <sup>-2</sup> s <sup>-1</sup> )	0.00945	0.00549	0.0258	0.0104
SLA (m <sup>2</sup> g C <sup>-1</sup> )	0.0133	0.0193	0.0142	0.0483
τ (year)	1.98	1.65	1.99	0.65
τ <sub>s</sub> (year)	0.0682	0.0583	0.0735	0.0102
C / N1	5.57	5.43	5	15.6
C / N2	77	20.2	10	52.7
γ1	6.25	37.2	20.8	39.6
γ2	5.26	35.5	27.3	5.58
γ3	0.257	0.471	0.361	0.000272
Inferred heteroscedastic error model inversions (HE2)				
g1	11.3	9.4	19.8	12.7
g0 (mol m <sup>-2</sup> s <sup>-1</sup> )	0.0276	0.00635	0.0298	0.0234
SLA (m <sup>2</sup> g C <sup>-1</sup> )	0.018	0.0158	0.0142	0.0797
τ (year)	1.69	1.8	1.27	0.822
τ <sub>s</sub> (year)	0.01	0.0892	0.0141	0.0104
C / N1	6.67	5.4	5.1	20.6
C / N2	22.9	15.5	14.9	10.1
γ1	7.83	21.7	20.6	38.5
γ2	6.14	21.9	19.1	9.79
γ3	0.503	0.896	0.145	0.505



**Figure 2.** Posterior distributions of the CARAIB model parameters sampled by the DREAM<sub>(ZS)</sub> algorithm with the inferred homoscedastic error model (HO2 inversions) for all sites. The default values (see Table 2) are depicted with a cross and the most likely values with a star. The  $x$  axes cover the whole prior ranges.



**Figure 3.** Measured and modelled GPP ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) (a), RECO ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) (b), ET ( $\text{mm day}^{-1}$ ) (c) and NEE ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) (d) at the Monte Bondone site for the inferred homoscedastic error model (inversion HO2). The ranges of the prediction uncertainty due to parameter uncertainty and the 95 % total predictive uncertainty (only for GPP, RECO and ET) are depicted by the dark and light grey shaded areas, respectively. Vertical arrows indicate the dates of the grass cutting.

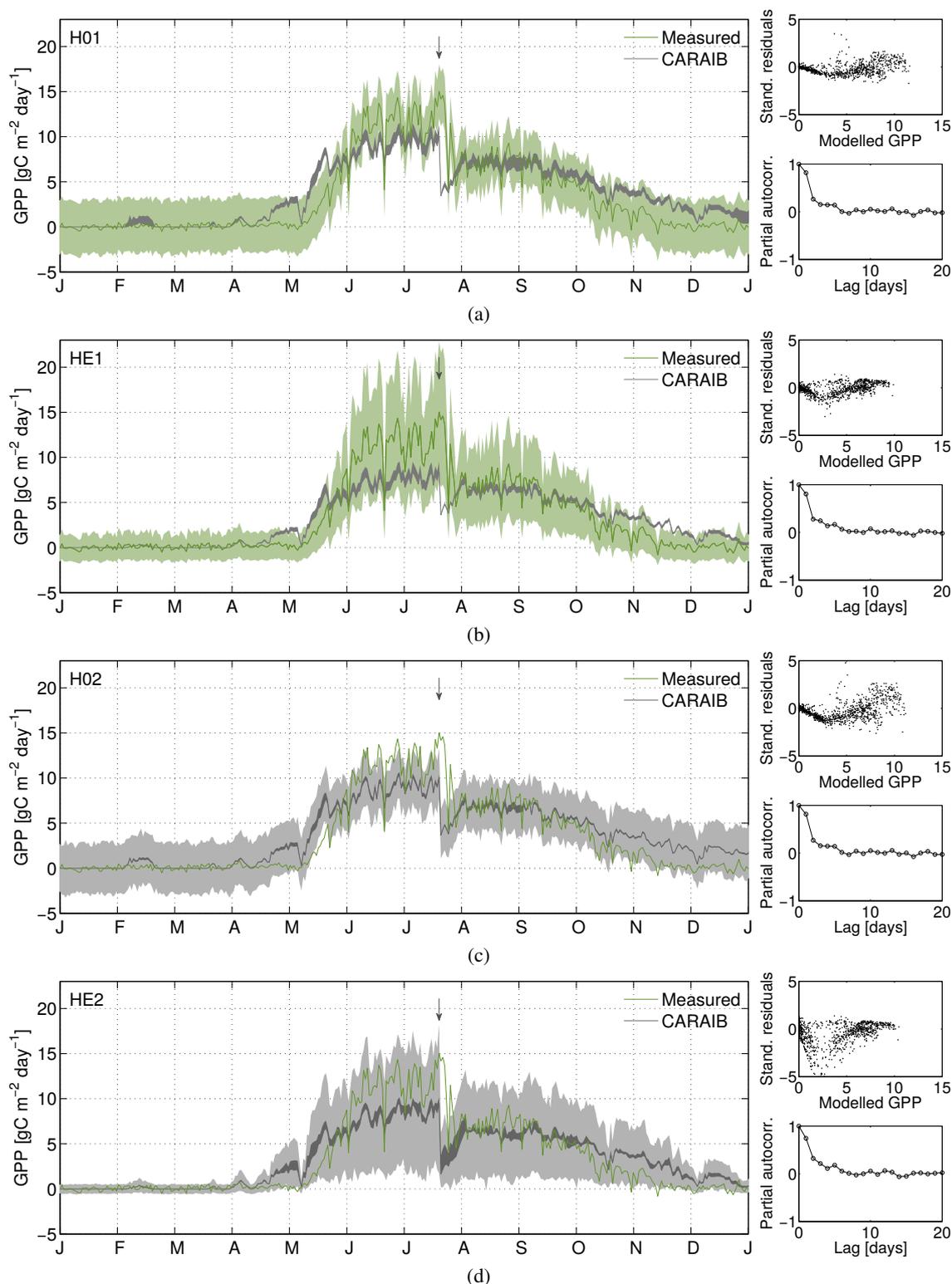


**Figure 4.** Measured and modelled GPP ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) for the Grillenburg (a), Oensingen (b) and Laqueuille (c) experimental sites. See Fig. 3 (a) for Monte Bondone. The ranges of the prediction uncertainty due to parameter uncertainty and the 95 % total predictive uncertainty are depicted by the dark and light grey shaded areas, respectively. Vertical arrows indicate the dates of the grass cutting (Grillenburg and Oensingen) and horizontal arrows the periods of grazing (Laqueuille).

est model efficiencies  $E$  were found for the NEE, which is not surprising since these data were not accounted for in the model inversions. While the ml values were generally the highest for the heteroscedastic inversions HE2, RMSE appeared larger for these inversions.

### 3.2.3 Homoscedastic and heteroscedastic eddy covariance residual errors

Considering homoscedastic or heteroscedastic eddy covariance residual errors resulted in different sampling of parameter posterior distributions and, therefore, different posterior modelled signals. As an example, Fig. 5 shows the measured and modelled GPP with their posterior ranges for the site of Monte Bondone in 2004, for both homoscedastic (a,



**Figure 5.** Measured and modelled GPP ( $\text{gC m}^{-2} \text{day}^{-1}$ ) at the Monte Bondone site in 2004 for the fixed homoscedastic HO1 (a) and heteroscedastic HE1 (b), inferred homoscedastic HO2 (c) and heteroscedastic HE2 (d) inversions. The measured GPP is depicted with a constant (a) and variable (b) uncertainty range. For the HO2 and HE2 inversions, the 95 % total predictive uncertainty interval is depicted using the light grey shaded areas. Standardized residuals and partial autocorrelation of residuals of GPP over the full simulation period are depicted to the right of each graph.

**Table 4.** Comparison between measured and modelled signals using most likely parameter values. The “ml” variable is the maximum value of the log-likelihood function.

	Grillenbug			Oensingen			Monte Bondone			Laqueuille		
	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>
Fixed homoscedastic error model inversions (HO1)												
ml	−5560			−7402			−5248			−8284		
GPP (g C m <sup>−2</sup> day <sup>−1</sup> )	1.797	0.726	0.791	2.231	0.600	0.757	1.742	0.755	0.831	2.151	0.521	0.751
RECO (g C m <sup>−2</sup> day <sup>−1</sup> )	1.498	0.502	0.695	1.269	0.772	0.803	1.036	0.832	0.878	1.529	0.688	0.743
ET (mm day <sup>−1</sup> )	0.623	0.309	0.565	0.670	0.612	0.758	0.500	0.784	0.849	1.128	0.144	0.474
NEE (g C m <sup>−2</sup> day <sup>−1</sup> )	1.774	−0.185	0.335	2.044	−0.115	0.449	1.424	−0.018	0.463	2.153	−0.382	0.219
Fixed heteroscedastic error model inversions (HE1)												
ml	−5324			−5961			−4879			−8078		
GPP (g C m <sup>−2</sup> day <sup>−1</sup> )	2.394	−0.018	0.706	2.405	0.353	0.767	1.932	0.585	0.814	2.679	0.001	0.695
RECO (g C m <sup>−2</sup> day <sup>−1</sup> )	1.977	−0.802	0.634	1.346	0.709	0.791	1.281	0.641	0.869	1.638	0.503	0.727
ET (mm day <sup>−1</sup> )	0.597	0.329	0.597	0.665	0.582	0.784	0.488	0.781	0.854	1.122	0.031	0.498
NEE (g C m <sup>−2</sup> day <sup>−1</sup> )	1.854	−1.491	0.198	2.086	−0.908	0.443	1.450	−0.501	0.429	2.138	−1.414	0.201
Inferred homoscedastic error model inversions (HO2)												
ml	−5161			−7074			−4550			−8321		
GPP (g C m <sup>−2</sup> day <sup>−1</sup> )	1.733	0.728	0.799	2.194	0.606	0.767	1.746	0.718	0.841	2.123	0.635	0.740
RECO (g C m <sup>−2</sup> day <sup>−1</sup> )	1.560	0.393	0.673	1.300	0.773	0.796	1.037	0.837	0.876	1.561	0.523	0.739
ET (mm day <sup>−1</sup> )	0.616	0.316	0.573	0.664	0.608	0.767	0.498	0.784	0.850	1.282	0.222	0.394
NEE (g C m <sup>−2</sup> day <sup>−1</sup> )	1.713	−0.139	0.367	2.034	−0.191	0.453	1.399	−0.332	0.478	2.052	−0.174	0.263
Inferred heteroscedastic error model inversions (HE2)												
ml	−4110			−6284			−3820			−7927		
GPP (g C m <sup>−2</sup> day <sup>−1</sup> )	1.929	0.669	0.744	2.306	0.467	0.762	1.875	0.645	0.811	2.225	0.475	0.737
RECO (g C m <sup>−2</sup> day <sup>−1</sup> )	1.751	0.344	0.582	1.350	0.758	0.781	1.244	0.661	0.869	1.674	0.621	0.702
ET (mm day <sup>−1</sup> )	0.574	0.403	0.629	0.663	0.589	0.781	0.492	0.784	0.852	1.283	0.221	0.393
NEE (g C m <sup>−2</sup> day <sup>−1</sup> )	1.652	0.002	0.384	2.071	−0.595	0.443	1.452	−0.246	0.433	2.217	−0.749	0.200

c) and heteroscedastic (b, d) cases. For the HO2 and HE2 inversions, the 95 % total predictive uncertainty is depicted using the light grey shaded areas. The measurement uncertainty is depicted only for fixed eddy covariance residual error inversions (a, b) for clarity. The measurement uncertainty is thus constant for the homoscedastic case (namely,  $\pm 3$  g C m<sup>−2</sup> day<sup>−1</sup> for HO1), while it varies linearly according to the GPP for the heteroscedastic case (HE1). These two options led to different behaviours of the modelled GPP using the posterior distributions, which better approached the high values of the measured data (in summer) in the homoscedastic cases and better fit the low values (in winter) in the heteroscedastic cases. Overall, in calibration, modelled signals with parameter values from the homoscedastic inversions were in a better agreement with the measured data than with the parameters from the heteroscedastic inversions. The same observation was also made for the other sites (not shown), as can also be observed in Table 4. However, the total predictive uncertainty range derived from the HE2 inversions was more consistent, as, e.g., it avoids unrealistic negative values of GPP. The standardized residuals, which were computed as the difference between measured

and modelled data divided by the standard deviation of the residual error, are depicted in Fig. 5 to the right of the GPP graphs. Heteroscedasticity of the GPP residual errors was fairly reduced but not fully removed by using the HE1 and HE2 heteroscedastic residual error models. Indeed, the standardized residuals still showed some small but complex heteroscedastic patterns. Partial autocorrelation of the residuals of the GPP was also depicted, and independence of the days of simulation from one another was reached after a few days.

### 3.2.4 Sampling of the standard deviation of the residual errors

Inversions with the sampling of the standard deviations of the residual errors resulted in posterior distributions of the standard deviation of the residual errors (HO2) and parameters of Eq. (7; HE2). Most likely values of these distributions (Table 5) depended on the experimental sites, being larger for Laqueuille and Oensingen, which can be related to the poorer agreement between measured and modelled data at these sites. Although the sampled standard deviations of the residual errors were lower than in the fixed inversions, there were no large differences between the inversions with fixed

**Table 5.** Most likely standard deviation of the residual errors (HO2) and parameters of Eq. (7; HE2).

	Grillenburg	Oensingen	Monte Bondone	Laqueuille
Inferred homoscedastic inversions (HO2)				
$\sigma_{\text{GPP}}$	1.81	2.29	1.79	2.22
$\sigma_{\text{RECO}}$	1.63	1.33	1.09	1.62
$\sigma_{\text{ET}}$	0.632	0.682	0.519	1.31
Inferred heteroscedastic inversions (HE2)				
$a_{\text{GPP}}$	0.211	0.65	0.336	1.09
$a_{\text{RECO}}$	0.12	0.334	0.162	0.514
$a_{\text{ET}}$	0.246	0.255	0.316	0.818
$b_{\text{GPP}}$	0.406	0.297	0.423	0.239
$b_{\text{RECO}}$	0.411	0.206	0.283	0.233
$b_{\text{ET}}$	0.273	0.255	0.12	0.175

model errors (HO1 & HE1) and inversions with inferred model errors (HO2 & HE2), neither in terms of agreement between measured and modelled signals (see Fig. 5 and Table 4) nor in the parameter posterior distributions (Table 3).

### 3.3 Model validation

Parameter values from the posterior distributions were tested for validation using eddy covariance data over different periods (for validation data sets, see Table 1). Figure 6 shows measured and modelled GPP values over the periods of calibration and validation in Monte Bondone. Not surprisingly, worse agreement between measured and modelled data is observed than in the calibration period. However, it is observed that the modelled GPP in validation in the HE2 inversions follows better the measured signal than in the HO2 inversions. Strikingly, at all the sites, the posterior parameter distributions derived from using the HE1 and HE2 heteroscedastic models are found to induce a better model performance in validation compared to the posterior distributions associated with the use of the homoscedastic models (Table 6). The difference between calibration and validation thus appeared smaller when using most likely parameter values from heteroscedastic inversions as compared to homoscedastic inversions. Among the different grassland sites, a similar performance pattern as for the calibration experiment is observed. Indeed, the Laqueuille site shows the worst performance statistics for each type of measurement data, whereas the Monte Bondone site overall presents the best fits to the data (Table 6).

## 4 Discussions

### 4.1 Measured and modelled signals

Bayesian inversions over the four grassland sites resulted in posterior distributions of parameters and posterior ranges of modelled signals (GPP, RECO, ET and NEE). Consider-

ing the inversion scenario HO2, there was, in general, good agreement between measured and modelled signals, with RMSEs ranging from 1.73 to 2.19  $\text{g C m}^{-2}\text{day}^{-1}$  and  $R^2$  being between 0.74 and 0.84 in terms of GPP. Using a dedicated model for soil organic carbon dynamics, De Bruijn et al. (2012) found an  $R^2$  of 0.68 for the modelling of the NEE at the Oensingen site over the same years. Comparing three large-scale lands surface models in simulating carbon fluxes over different ecosystems, Balzarolo et al. (2014) noticed that grassland and crop sites were more difficult to model compared to forest sites. Using data from 13 grassland sites over Europe, including Laqueuille and Grillenburg, they found average RMSEs between measured and modelled GPP ranging from 2.45 to 3.57  $\text{g C m}^{-2}\text{day}^{-1}$  and  $R^2$  from 0.37 to 0.56. These larger discrepancies compared to our study are mainly to be related to the fact that the large-scale models were used without site calibrations. Modelling of carbon fluxes was also performed at the Oensingen site over the same years in Calanca et al. (2007), using a dedicated grassland model, PaSim. In that study, no numerical comparison between measured and modelled data was computed at a daily resolution, but the relative departures between measured (eddy covariance) and modelled data were given by year of simulation and ranged from  $-11$  to  $-21$  % in terms of the annual sum of GPP. In our study, the annual relative departures in the annual sum of GPP in Oensingen ranged from 0.7 to 9 % with the calibration data set and up to 63 % with the validation data set. In a similar experiment involving the inversion of eddy covariance data from forest sites, Fox et al. (2009) found RMSEs between measured and modelled NEE of 0.7 and 1.3  $\text{g C m}^{-2}\text{day}^{-1}$  for two different sites in calibration and of 1.5  $\text{g C m}^{-2}\text{day}^{-1}$  in validation. These values are lower than in our study, but the measured NEE data were not used in the model inversion here, unlike in the inversions in Fox et al. (2009).

It could be observed that measured eddy covariance data have stronger dynamics than the modelled signals, that is, modelled signals could not follow the fast fluctuations of the measured signals and, in particular, simulate high GPP values. This could be related to the different time resolutions between the model and data. The CARAIB model is based on meteorological data averaged daily. However, photosynthesis and respiration processes are computed at a 2-hour time step before being aggregated to a daily resolution, and the model assumes a symmetry with respect to solar noon time (Otto et al., 2002) to save computation resource. Moreover, in the CARAIB model, solar fluxes are calculated assuming a constant cloudiness over the day and temperature is varied using a sinusoidal function between the minimal and maximal temperatures, which were fixed at midnight and noon, respectively. These shortcomings were necessary to save computation resources and to account for data scarcity in global vegetation modelling. Eddy covariance data, however, are typically acquired at a time frequency of 5 or 10 Hz (Aubinet et al., 2012) and can thus capture high-frequency

**Table 6.** Validation of the calibrated model using most likely parameter values from the inversions.

	Grillenburg			Oensingen			Monte Bondone			Laqueuille		
	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>	RMSE	<i>E</i>	<i>R</i> <sup>2</sup>
Fixed homoscedastic error model inversions (HO1)												
GPP (g C m <sup>-2</sup> day <sup>-1</sup> )	2.467	0.581	0.694	3.000	0.523	0.576	2.234	0.711	0.803	4.160	-0.307	0.690
RECO (g C m <sup>-2</sup> day <sup>-1</sup> )	1.284	0.768	0.799	1.560	0.732	0.747	1.389	0.733	0.871	4.444	-2.363	0.607
ET (mm day <sup>-1</sup> )	0.642	0.520	0.602	0.732	0.662	0.700	0.504	0.839	0.848	1.198	0.226	0.382
NEE (g C m <sup>-2</sup> day <sup>-1</sup> )	1.880	0.197	0.453	2.332	0.049	0.217	1.526	0.446	0.475	2.481	-0.176	0.171
Fixed heteroscedastic error model inversions (HE1)												
GPP (g C m <sup>-2</sup> day <sup>-1</sup> )	2.030	0.716	0.748	2.803	0.584	0.585	1.812	0.810	0.832	2.707	0.446	0.679
RECO (g C m <sup>-2</sup> day <sup>-1</sup> )	1.221	0.790	0.871	1.535	0.740	0.747	0.960	0.873	0.881	3.101	-0.638	0.595
ET (mm day <sup>-1</sup> )	0.610	0.567	0.613	0.678	0.709	0.714	0.502	0.840	0.851	1.142	0.296	0.393
NEE (g C m <sup>-2</sup> day <sup>-1</sup> )	1.972	0.117	0.281	2.206	0.149	0.194	1.415	0.524	0.536	2.339	-0.045	0.180
Inferred homoscedastic error model inversions (HO2)												
GPP (g C m <sup>-2</sup> day <sup>-1</sup> )	2.651	0.516	0.663	3.008	0.520	0.571	2.315	0.690	0.765	2.730	0.437	0.705
RECO (g C m <sup>-2</sup> day <sup>-1</sup> )	1.335	0.749	0.777	1.614	0.713	0.735	1.398	0.730	0.849	2.050	0.284	0.646
ET (mm day <sup>-1</sup> )	0.639	0.525	0.603	0.710	0.682	0.705	0.502	0.841	0.850	1.364	-0.003	0.353
NEE (g C m <sup>-2</sup> day <sup>-1</sup> )	2.010	0.082	0.451	2.337	0.045	0.203	1.598	0.393	0.414	2.047	0.200	0.309
Inferred heteroscedastic error model inversions (HE2)												
GPP (g C m <sup>-2</sup> day <sup>-1</sup> )	2.361	0.617	0.699	2.825	0.577	0.588	1.805	0.811	0.830	2.294	0.603	0.701
RECO (g C m <sup>-2</sup> day <sup>-1</sup> )	1.220	0.790	0.815	1.549	0.735	0.749	0.974	0.869	0.883	2.472	-0.041	0.625
ET (mm day <sup>-1</sup> )	0.617	0.557	0.614	0.683	0.705	0.713	0.501	0.841	0.850	1.370	-0.013	0.351
NEE (g C m <sup>-2</sup> day <sup>-1</sup> )	1.837	0.233	0.364	2.227	0.132	0.199	1.398	0.535	0.547	2.240	0.042	0.179

fluxes. Even though eddy covariance data were aggregated over time to a daily time resolution, the high-frequency acquisition rate ensures that effects of abrupt meteorological events are recorded. Increasing the time resolution of the CARAIB model would help to better simulate ecophysiological processes at a high frequency. Alternatively, a simple workaround to deal with the different time dynamics would be to apply a filter based on a moving window of some days in order to smooth measured (and modelled) eddy covariance data before computing the statistical indicators, as done in Calanca et al. (2007).

Another modelling limitation is that model parameters are assumed as constant along the season, although plants traits are known to evolve throughout the season and plants acclimate to specific climate conditions. As a result, the effect of similar climatic conditions does not necessary result in similar eddy covariance measurements.

In general, there was poorer agreement between measured and modelled signals (GPP, RECO, ET and NEE) in Laqueuille than at the other experimental sites. This poorer agreement can probably be related to the grazing instead of the cutting that occurs in Laqueuille. Grazing was more difficult to simulate because of the expert-knowledge conversion between the given cattle charge and the biomass removal. As a result, grass cutting is better constrained in the model compared to grazing, as was already shown in the Laqueuille ex-

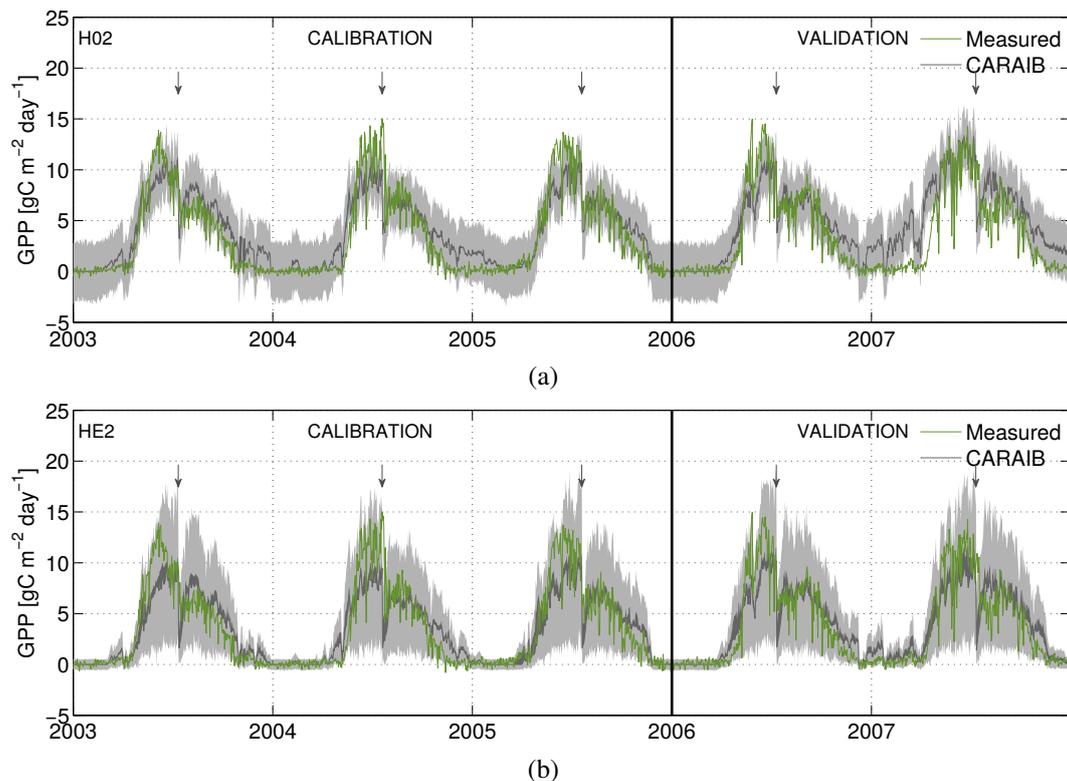
perimental site by Calanca et al. (2007), who, however, used the grassland model PaSim.

All the same, besides the average statistical indicators between measured and modelled signals, the performance of the calibration might be also evaluated against specific scientific or operational objectives. For instance, the accurate modelling of the grass cutting or the computation of annual budgets of carbon in grassland (e.g., Soussana et al., 2007) might show different performances, depending on the timescale on which the processes are analysed.

## 4.2 Eddy covariance residual errors

### 4.2.1 Homoscedastic and heteroscedastic eddy covariance residual errors

Bayesian inversions were conducted considering homoscedasticity and heteroscedasticity in the eddy covariance residual errors. Figure 5 showed that accounting for heteroscedasticity in eddy covariance residual errors permitted a better simulation of low-magnitude signals (winter), but at the same time, it penalized the modelling of high-magnitude signals (summer). Actually, it is worth remarking that, in carrying out inversions considering heteroscedastic measurement errors, we do not attempt to produce smaller RMSEs between measured and modelled data compared to homoscedastic scenarios since larger errors are considered



**Figure 6.** Measured and modelled GPP ( $\text{g C m}^{-2} \text{ day}^{-1}$ ) at the Monte Bondone site in calibration (2003–2005) and validation (2006–2007) for the inferred homoscedastic HO2 (a) and heteroscedastic HE2 (b) inversions. The 95 % confidence interval of total predictive uncertainty is depicted using the light grey shaded areas.

for high peaks in the signals. However, in validation, the posterior parameter distributions derived from using the heteroscedastic residual error models outperform their counterparts derived from using the homoscedastic residual error models. This important finding reveals that, despite inducing larger RMSE values in calibration, the use of a heteroscedastic residual error model leads to a more robust parameter estimation.

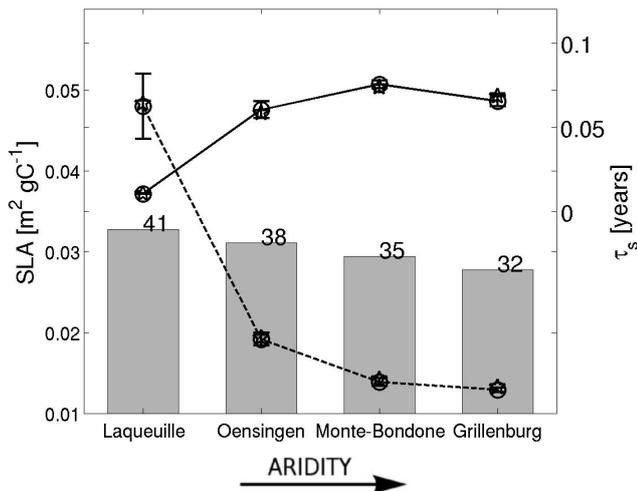
Since eddy covariance data are known to show heteroscedasticity, accounting for a heteroscedastic model of the residual errors in the inversions is more conceptually sound for ensuring unbiased parameter posterior distributions. However, we showed that considering a linear heteroscedastic model of the residual errors only partly removed heteroscedasticity in the standardized residual values (Fig. 5b and d). Other kinds of heteroscedastic models (i.e., non-linear) might be tested, but the residual distributions did not show any clear trend for all sites.

It is also worth noting that a substantial fraction of the large residual errors is caused by the tendency of the CARAIB model to underestimate the observed GPP summer peaks. As discussed above, this is related to a slower temporal resolution of the model compared to that of the measured data. To overcome this model inadequacy, further model modifications are necessary to increase the time reso-

lution of the model. Another model improvement would be to simulate varying model parameter values as a function of the time of the year, since plant traits actually evolve over the course of the seasons. However, this would come at the cost of a large increase in model complexity.

#### 4.2.2 Sampling of the standard deviation of residual errors

Sampling the standard deviation of the residual errors, i.e., the inversions HO2 and HE2, resulted in similar parameter samplings and modelling as the inversions HO1 and HE1, respectively. Some performance criteria were better with the sampling of the residual standard deviations, while others were not. As expected, the most likely standard deviations of the residuals errors were close to the RMSE obtained in the inversions HO2. The benefit of these values is that they inform us about the level of the uncertainties in the eddy covariance data with respect to the model used to invert the data, e.g., uncertainties in GPP ranged from 1.79 to 2.29  $\text{g C m}^{-2} \text{ day}^{-1}$ , in RECO from 1.09 to 1.63  $\text{g C m}^{-2} \text{ day}^{-1}$  and in ET from 0.52 to 1.31  $\text{mm day}^{-1}$ . They could be used to weight different eddy covariance data in multi-objective inverse modelling.



**Figure 7.** Posterior distributions of the specific leaf area (SLA, dashed line) and characteristic mortality time in stress conditions  $\tau_s$  (plain line) for the four sites (HO2 inversions values), classified as a function of increasing aridity by the De Martonne–Gottman index (grey bars). The mean of the posterior distributions and the most likely parameter values are depicted with a circle and a star, respectively. The error bars stand for one standard deviation around the mean.

### 4.3 Parameter values across sites

Posterior distributions of parameters showed contrasting values that could be linked to the characteristics of the experimental sites. For instance, the specific leaf area (SLA) is known to depend on many factors (Marcelis et al., 1998), such as leaf age, temperature, light intensity, aridity and soil nutrient content. Thick leaves (low SLA) are more adapted to dry ecosystems due to their greater capacity to retain water. Although none of the four grassland sites are strictly characterized by a dry climate, it is interesting to note that the posterior parameter distributions for SLA were negatively correlated with the aridity, inversely expressed by the De Martonne–Gottman index (Fig. 7), that is, SLA decreases with increasing aridity. The largest SLA (thin leaves) were found for Laqueuille, which can be related to the permanent grazing that constantly regenerates young leaves, since young leaves are characterized by high SLA. The large SLA values in Oensingen can be related to more intensive management conditions (fertilization, more frequent cuts).

Contrarily to SLA, the characteristic mortality time in stress conditions  $\tau_s$  appeared to be positively correlated with the site aridity (Fig. 7). A larger  $\tau_s$  value means a larger water stress resistance for the plants in Grillenburg and Monte Bondone.

The values of  $g_1$  were drastically different between Oensingen and the three other sites (Table 3). In addition, for these three sites, the values appeared much higher compared to the default values ( $g_1 = 9$ ) and other values com-

monly encountered in the literature (Van Wijk et al., 2000; Medlyn et al., 2011). It is known that  $g_1$  should increase with humid conditions and temperature (Medlyn et al., 2011), as it is positively related to the marginal water cost of carbon gain. However, the high values of  $g_1$  here could not really be related to a warmer or wetter climate as compared to Oensingen. A possible explanation could be related to the different dynamics of the model and the measurements, as already explained herein before. As the model cannot simulate the high GPP values that are observed in the eddy covariance data, the Bayesian algorithm could have compensated for this by sampling high values of  $g_1$  that increase stomatal conductance.

More broadly, ecophysiological differences between the grassland sites resulted in parameter posterior distributions that can be either drastically different or common between the sites (Fig. 2). If it appears that site-specific parameter values are needed, it means that the model has to be refined by accounting for the ecophysiological dependence of the parameters. If not, generalized parameter values could be used, meaning that they are independent of the site on which they were determined or even independent of the plant species, as recently claimed by Yuan et al. (2014). Determining a common set of the parameter distributions among the four sites could be done either by (1) merging the four posterior distributions after independent inversions of the data of each site or (2) merging the eddy covariance data of the four sites in one single MCMC sampling, as discussed in Kuppel et al. (2012).

## 5 Conclusions

Bayesian inversions of the CARAIB dynamic vegetation model were performed using eddy covariance data (GPP, RECO, ET) at four experimental grassland sites. A specific version of the CARAIB model was developed for this application, with functions related to grassland management, i.e., grass cutting and grazing. Posterior parameter and predictive distributions were compared for different statistical models of the eddy covariance residual errors: (1) assuming homoscedasticity or heteroscedasticity of the residual errors and (2) fixing beforehand or jointly inferring the variances of the residual errors. There was, in general, good agreement between measured and modelled signals for the calibration data sets with RMSEs of daily gross primary productivity (GPP), ecosystem respiration (RECO) and evapotranspiration (ET) ranging from 1.73 to 2.19, 1.04 to 1.56 g C m<sup>-2</sup> day<sup>-1</sup> and 0.50 to 1.28 mm day<sup>-1</sup>, respectively. Since the four sites belong to a long-standing network of eddy covariance data measurements, comparisons with previous studies could be made.

Although the eddy covariance measurement errors are known to be heteroscedastic, the use of a homoscedastic error model led to a better model performance in calibration compared to using a heteroscedastic error model. Nevertheless,

a model validation experiment revealed that CARAIB models calibrated by means of a heteroscedastic error model outperform those calibrated assuming homoscedastic residual errors. Posterior parameter distributions derived from using a heteroscedastic model of the residuals are therefore more sound and robust, even though heteroscedasticity could not be fully removed. Therefore, our results support the use of a heteroscedastic residual error model for inverting eddy covariance data and inferring posterior parameter distributions.

Systematic model–data discrepancies were also found for the largest observed GPP values. This can be attributed to the low temporal resolution of the photosynthetic processes in the CARAIB model, among other model inadequacies. Modelling performance varied among the four sites, with poorer performances at Laqueuille because of the greater difficulty of modelling grazing compared to grass cutting. Lastly, site-specific posterior parameter distributions obtained for the four grasslands were compared and discussed with respect to grassland characteristics. Specific leaf area and characteristic mortality time parameters appeared to be related to site aridity.

*Acknowledgements.* This research was funded by the “Direction Générale Opérationnelle de l’Économie, de l’Emploi & de la Recherche” (DGO6), Wallonie, Belgium. This study was made within the framework of the FACCE/MACSUR knowledge hub, a pan-European collaborative project based on the modelling of agriculture systems facing climate change. In particular, this study formed part of the intercomparison of grassland models task, led by Gianni Bellocchi, that aims to compare and improve the performance of eight models in grassland growth simulations.

Edited by: T. Keenan

## References

- Ammann, C., Flechard, C. R., Leifeld, J., Neftel, A., and Fuhrer, J.: The carbon budget of newly established temperate grassland depends on management intensity, *Agr. Ecosyst. Environ.*, 121, 5–20, doi:10.1016/j.agee.2006.12.002, 2007.
- Aubinet, M., Vesala, T., and Papale, D. (Eds.): *Eddy Covariance: A Practical Guide to Measurement and Data Analysis*, Springer, 173–209, 2012.
- Ball, Woodrow, I., and Berry, J.: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in: *Progress in Photosynthesis Research*, edited by: Biggins, J., 221–224, Springer Netherlands, doi:10.1007/978-94-017-0519-6\_48, 1987.
- Balzarolo, M., Boussetta, S., Balsamo, G., Beljaars, A., Maignan, F., Calvet, J.-C., Lafont, S., Barbu, A., Poulter, B., Chevalier, F., Szczypta, C., and Papale, D.: Evaluating the potential of large-scale simulations to predict carbon fluxes of terrestrial ecosystems over a European Eddy Covariance network, *Biogeosciences*, 11, 2661–2678, doi:10.5194/bg-11-2661-2014, 2014.
- Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, *Glob. Change Biol.*, 13, 679–706, doi:10.1111/j.1365-2486.2006.01305.x, 2007.
- Braswell, B. H., Sacks, W. J., Linder, E., and Schimel, D. S.: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, *Glob. Change Biol.*, 11, 335–355, doi:10.1111/j.1365-2486.2005.00897.x, 2005.
- Calanca, P., Vuichard, N., Campbell, C., Viovy, N., Cozic, A., Fuhrer, J., and Soussana, J. F.: Simulating the fluxes of CO<sub>2</sub> and N<sub>2</sub>O in European grasslands with the Pasture Simulation Model (PaSim), *Agr. Ecosys. Environ.*, 121, 164–174, doi:10.1016/j.agee.2006.12.010, 2007.
- De Bruijn, A. M. G., Calanca, P., Ammann, C., and Fuhrer, J.: Differential long-term effects of climate change and management on stocks and distribution of soil organic carbon in productive grasslands, *Biogeosciences*, 9, 1997–2012, doi:10.5194/bg-9-1997-2012, 2012.
- Dumont, B., Leemans, V., Mansouri, M., Bodson, B., Destain, J. P., and Destain, M. F.: Parameter identification of the STICS crop model, using an accelerated formal MCMC approach, *Environ. Modell. Softw.*, 52, 121–135, doi:10.1016/j.envsoft.2013.10.022, 2014.
- Evans, J. R. and Poorter, H.: Photosynthetic acclimation of plants to growth irradiance: the relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain, *Plant Cell Environ.*, 24, 755–767, doi:10.1046/j.1365-3040.2001.00724.x, 2001.
- FAO STAT: <http://faostat.fao.org/site/377/DesktopDefault.aspx> (last access: 28 January 2015), 2011.
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species, *Planta*, 149, 78–90, doi:10.1007/bf00386231, 1980.
- Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, *Agr. Forest Meteorol.*, 149, 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.
- Friend, A. D., Arneeth, A., Kiang, N. Y., Lomas, M., Ogée, J., Rördenbeck, C., Running, S. W., Santaren, J.-D., Sitch, S., Viovy, N., Ian Woodward, F., and Zaehle, S.: FLUXNET and modelling the global carbon cycle, *Glob. Change Biol.*, 13, 610–633, doi:10.1111/j.1365-2486.2006.01223.x, 2007.
- Gelman, A. and Rubin, D. B.: Inference from Iterative Simulation Using Multiple Sequences, *Stat. Sci.*, 7, 457–472, doi:10.1214/ss/1177011136, 1992.
- Gottschalk, P., Wattenbach, M., Neftel, A., Fuhrer, J., Jones, M., Lanigan, G., Davis, P., Campbell, C., Soussana, J. F., and Smith, P.: The role of measurement uncertainties for the simulation of grassland net ecosystem exchange (NEE) in Europe, *Agr. Ecosyst. Environ.*, 121, 175–185, doi:10.1016/j.agee.2006.12.026, 2007.
- Graux, A.-I., Bellocchi, G., Lardy, R., and Soussana, J.-F.: Ensemble modelling of climate change risks and opportunities for managed grasslands in France, *Agr. Forest Meteorol.*, 170, 114–131, doi:10.1016/j.agrformet.2012.06.010, 2013.

- Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O'Hara, R. B., Scheiter, S., and Huth, A.: Connecting dynamic vegetation models to data – an inverse perspective, *J. Biogeogr.*, 39, 2240–2252, doi:10.1111/j.1365-2699.2012.02745.x, 2012.
- Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, *Tree Physiol.*, 25, 873–885, doi:10.1093/treephys/25.7.873, 2005.
- Klump, K., Tallec, T., Guix, N., and Soussana, J.-F.: Long-term impacts of agricultural practices and climatic variability on carbon storage in a permanent pasture, *Glob. Change Biol.*, 17, 3534–3545, doi:10.1111/j.1365-2486.2011.02490.x, 2011.
- Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, *Glob. Change Biol.*, 11, 1333–1351, doi:10.1111/j.1365-2486.2005.00977.x, 2005.
- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosciences*, 9, 3757–3776, doi:10.5194/bg-9-3757-2012, 2012.
- Laloy, E. and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing, *Water Resour. Res.*, 48, W01526, doi:10.1029/2011wr010608, 2012.
- Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, *Biogeosciences*, 5, 1311–1324, doi:10.5194/bg-5-1311-2008, 2008.
- Laurent, J., François, L., Barhen, A., Bel, L., and Cheddadi, R.: European bioclimatic affinity groups: Data-model comparisons, *Global Planet. Change*, 61, 28–40, doi:10.1016/j.gloplacha.2007.08.017, 2008.
- Laurent, J. M., Bar-Hen, A., François, L., Ghislain, M., Cheddadi, R., and Wiser, S.: Refining vegetation simulation models: From plant functional types to bioclimatic affinity groups of plants, *J. Veg. Sci.*, 15, 739–746, doi:10.1658/1100-9233(2004)015%5B0739:rvmfp%5D2.0.co;2, 2004.
- Leuning, R.: A critical appraisal of a combined stomatal-photosynthesis model for C3 plants, *Plant Cell Environ.*, 18, 339–355, doi:10.1111/j.1365-3040.1995.tb00370.x, 1995.
- Ma, S., Acutis, M., Barcza, Z., Ben Touhami, H., Doro, L., Hidy, D., Koechy, M., Minet, J., Lellei-Kovacs, E., Perego, A., Rolinski, S., Ruget, F., Seddaiu, G., Wu, L., and Bellocchi, G.: The grassland model intercomparison of the MACSUR (Modelling European Agriculture with Climate Change for Food Security) European knowledge hub, in: Proceedings of the 7th International Congress on Environmental Modelling and Software (iEMSs) June 15–19, 2014, San Diego, California, USA, 61, 2014.
- Marcelis, L. F. M., Heuvelink, E., and Goudriaan, J.: Modelling biomass production and yield of horticultural crops: a review, *Sci. Hortic.*, 74, 83–111, doi:10.1016/s0304-4238(98)00083-1, 1998.
- Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., and Wingate, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance, *Glob. Change Biol.*, 17, 2134–2144, doi:10.1111/j.1365-2486.2010.02375.x, 2011.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.
- Nemry, B., François, L., Warnant, P., Robinet, F., and Gérard, J. C.: The seasonality of the CO<sub>2</sub> exchange between the atmosphere and the land biosphere: A study with a global mechanistic vegetation model, *J. Geophys. Res.*, 101, 7111–7125, doi:10.1029/95jd03656, 1996.
- O'Mara, F. P.: The role of grasslands in food security and climate change., *Annals of botany*, 110, 1263–1270, doi:10.1093/aob/mcs209, 2012.
- Otto, D., Rasse, D., Kaplan, J., Warnant, P., and François, L.: Biospheric carbon stocks reconstructed at the Last Glacial Maximum: comparison between general circulation models using prescribed and computed sea surface temperatures, *Global Planet. Change*, 33, 117–138, doi:10.1016/s0921-8181(02)00066-8, 2002.
- Patenaude, G., Milne, R., Van Oijen, M., Rowland, C. S., and Hill, R. A.: Integrating remote sensing datasets into ecological modelling: a Bayesian approach, *Int. J. Remote Sens.*, 29, 1295–1315, doi:10.1080/01431160701736414, 2008.
- Poorter, H. and De Jong, R.: A comparison of specific leaf area, chemical composition and leaf construction costs of field plants from 15 habitats differing in productivity, *New Phytol.*, 143, 163–176, doi:10.1046/j.1469-8137.1999.00428.x, 1999.
- Prescher, A.-K., Grünwald, T., and Bernhofer, C.: Land use regulates carbon budgets in eastern Germany: From NEE to NBP, *Agr. Forest Meteorol.*, 150, 1016–1025, doi:10.1016/j.agrformet.2010.03.008, 2010.
- Reichert, P. and Schuwirth, N.: Linking statistical bias description to multiobjective model calibration, *Water Resour. Res.*, 48, W09543, doi:10.1029/2011wr011391, 2012.
- Richardson, A. D. and Hollinger, D. Y.: Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models, *Agr. Forest Meteorol.*, 131, 191–208, doi:10.1016/j.agrformet.2005.05.008, 2005.
- Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Mofat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO<sub>2</sub> flux measurement uncertainty inferred from model residuals, *Agr. Forest Meteorol.*, 148, 38–50, doi:10.1016/j.agrformet.2007.09.001, 2008.
- Rivington, M., Matthews, K. B., Bellocchi, G., and Buchan, K.: Evaluating uncertainty introduced to process-based simulation model estimates by alternative sources of meteorological data, *Ag. Syst.*, 88, 451–471, doi:10.1016/j.agsy.2005.07.004, 2006.
- Robert, C. and Casella, G.: Monte Carlo statistical methods., Springer texts in statistics, Springer, 580 pp., 2004.
- Ruget, M.: Modeling Tiller Density, Growth, and Yield of Mediterranean Perennial Grasslands with STICS, *Crop science*, 49, 2379–2385, 2009.
- Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, *Glob. Change Biol.*, 9, 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

- Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., Betts, R., Ciais, P., Cox, P., Friedlingstein, P., Jones, C. D., Prentice, I. C., and Woodward, F. I.: Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs), *Glob. Change Biol.*, 14, 2015–2039, doi:10.1111/j.1365-2486.2008.01626.x, 2008.
- Soussana, J. F., Loiseau, P., Vuichard, N., Ceschia, E., Balesdent, J., Chevallier, T., and Arrouays, D.: Carbon cycling and sequestration opportunities in temperate grasslands, *Soil Use Manage.*, 20, 219–230, doi:10.1111/j.1475-2743.2004.tb00362.x, 2004.
- Soussana, J. F., Allard, V., Pilegaard, K., Ambus, P., Amman, C., Campbell, C., Ceschia, E., Clifton-Brown, J., Czobel, S., Domingues, R., Flechard, C., Fuhrer, J., Hensen, A., Horvath, L., Jones, M., Kasper, G., Martin, C., Nagy, Z., Neftel, A., Raschi, A., Baronti, S., Rees, R. M., Skiba, U., Stefani, P., Manca, G., Sutton, M., Tuba, Z., and Valentini, R.: Full accounting of the greenhouse gas (CO<sub>2</sub>, N<sub>2</sub>O, CH<sub>4</sub>) budget of nine European grassland sites, *Agr. Ecosyst. Environ.*, 121, 121–134, doi:10.1016/j.agee.2006.12.022, 2007.
- ter Braak, C. and Vrugt, J.: Differential Evolution Markov Chain with snooker updater and fewer chains, *Stat. Comput.*, 18, 435–446, doi:10.1007/s11222-008-9104-9, 2008.
- Van Wijk, M. T., Dekker, S. C., Bouten, W., Bosveld, F. C., Kohsiek, W., Kramer, K., and Mohren, G. M. J.: Modeling daily gas exchange of a Douglas-fir forest: comparison of three stomatal conductance models with and without a soil water stress function, *Tree Physiol.*, 20, 115–122, doi:10.1093/treephys/20.2.115, 2000.
- Vile, D., Garnier, E., Shipley, B., Laurent, G., Navas, M.-L., Roumet, C., Lavorel, S., Díaz, S., Hodgson, J. G., Lloret, F., Midgley, G. F., Poorter, H., Rutherford, M. C., Wilson, P. J., and Wright, I. J.: Specific Leaf Area and Dry Matter Content Estimate Thickness in Laminar Leaves, *Ann.Bot.*, 96, 1129–1136, doi:10.1093/aob/mci264, 2005.
- Vrugt, J. A., Braak, C. J. F. T., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlin. Sci. Num.*, 10, 273–290, 2009.
- Warnant, P.: Modélisation du cycle du carbone dans la biosphère continentale à l'échelle globale, Ph.D. thesis, Université de Liège, Liège, 118 pp., 1999.
- Warnant, P., François, L., Strivay, D., and Gérard, J. C.: CARAIB: A global model of terrestrial biological productivity, *Global Biogeochem. Cycles*, 8, 255–270, doi:10.1029/94gb00850, 1994.
- Williams, J. R., Arnold, J. G., Kiniry, J. R., Gassman, P. W., and Green, C. H.: History of model development at Temple, Texas, *Hydrolog. Sci. J.*, 53, 948–960, doi:10.1623/hysj.53.5.948, 2008.
- Wilson, P. J., Thompson, K., and Hodgson, J. G.: Specific leaf area and leaf dry matter content as alternative predictors of plant strategies, *New Phytol.*, 143, 155–162, doi:10.1046/j.1469-8137.1999.00427.x, 1999.
- Wohlfahrt, G., Anderson-Dunn, M., Bahn, M., Balzarolo, M., Berninger, F., Campbell, C., Carrara, A., Cescatti, A., Christensen, T., Dore, S., Eugster, W., Friborg, T., Furger, M., Gianelle, D., Gimeno, C., Hargreaves, K., Hari, P., Haslwanter, A., Johansson, T., Marcolla, B., Milford, C., Nagy, Z., Nemitz, E., Rogiers, N., Sanz, M., Siegwolf, R., Susiluoto, S., Sutton, M., Tuba, Z., Ugolini, F., Valentini, R., Zorer, R., and Cernusca, A.: Biotic, Abiotic, and Management Controls on the Net Ecosystem CO<sub>2</sub> Exchange of European Mountain Grassland Ecosystems, *Ecosystems*, 11, 1338–1351, doi:10.1007/s10021-008-9196-2, 2008.
- Wu, L., McGechan, M. B., McRoberts, N., Baddeley, J. A., and Watson, C. A.: SPACSYS: Integration of a 3D root architecture component to carbon, nitrogen and water cycling-Model description, *Ecol. Modell.*, 200, 343–359, doi:10.1016/j.ecolmodel.2006.08.010, 2007.
- Yuan, W., Cai, W., Liu, S., Dong, W., Chen, J., Arain, M. A., Blanken, P. D., Cescatti, A., Wohlfahrt, G., Georgiadis, T., Genesio, L., Gianelle, D., Grelle, A., Kiely, G., Knohl, A., Liu, D., Marek, M. V., Merbold, L., Montagnani, L., Panferov, O., Peltoniemi, M., Rambal, S., Raschi, A., Varlagin, A., and Xia, J.: Vegetation-specific model parameters are not required for estimating gross primary production, *Ecol. Modell.*, 292, 1–10, doi:10.1016/j.ecolmodel.2014.08.017, 2014.
- Zhao, Y., Ciais, P., Peylin, P., Viovy, N., Longdoz, B., Bonnefond, J. M., Rambal, S., Klumpp, K., Olioso, A., Cellier, P., Maignan, F., Eglin, T., and Calvet, J. C.: How errors on meteorological variables impact simulated ecosystem fluxes: a case study for six French sites, *Biogeosciences*, 9, 2537–2564, doi:10.5194/bg-9-2537-2012, 2012.