Geoscientific
Model Development

Open Access

Global sensitivity analysis, probabilistic calibration, and predictive assessment for the data assimilation linked ecosystem carbon model

C. Safta[1], D. M. Ricciuto[2], K. Sargsyan[1], B. Debusschere[1], H. N. Najm[1], M. Williams[3], and P. E. Thornton[2]

[1]Sandia National Labs, Livermore, CA 94551, USA
[2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[3]School of GeoSciences and National Centre for Earth Observation, University of Edinburgh, EH9 EJN, UK

Correspondence to: C. Safta (csafta@sandia.gov)

Abstract. In this paper we propose a probabilistic framework for an uncertainty quantification (UQ) study of a carbon cycle model and focus on the comparison between steady-state and transient simulation setups. A global sensitivity analysis (GSA) study indicates the parameters and parameter couplings that are important at different times of the year for quantities of interest (QoIs) obtained with the data assimilation linked ecosystem carbon (DALEC) model. We then employ a Bayesian approach and a statistical model error term to calibrate the parameters of DALEC using net ecosystem exchange (NEE) observations at the Harvard Forest site. The calibration results are employed in the second part of the paper to assess the predictive skill of the model via posterior predictive checks.

1 Introduction

Climate studies strongly depend on the modeling of the carbon cycle. Carbon cycle models, in turn, strongly depend on the capability of current land models to simulate the terrestrial ecosystem and to capture carbon exchanges between land and atmosphere. There have been a significant number of studies looking to leverage the increasing number of experimental observations and calibrate parameters in several terrestrial ecosystem models. These studies have faced a number of challenges related to handling data and measurement errors from multiple sources, formalizing model error, and dealing with parameter observability and data sparsity to name a few. In this paper we propose a probabilistic framework to estimate parameters for a process-based ecosystem

model. Representative studies, both probabilistic and non-probabilistic, are reviewed below.

Over the past 2 decades several studies employed data assimilation techniques to calibrate carbon cycle models. Here we briefly discuss the works that motivated the current study. Kaminski et al. (2002, 2012) used an adjoint approach to infer model parameters for a terrestrial biosphere model based on observational data streams. The variational data assimilation problem was formulated based on Bayes' theorem with both the likelihood and the prior presumed Gaussian. It was found that models employing optimized parameters show clear improvements when checked against independent observations compared to non-optimized parameters. Similar approaches were employed by Rayner et al. (2005), Tjiputra et al. (2007), and Kuppel et al. (2012) to estimate parameters of ecosystem models.

Some of the above studies start from a Bayesian framework when setting the cost function for a least-square fitting procedure. The resulting probability densities for model parameters are approximated as multivariate Gaussian densities near the maximum a posteriori (MAP) estimate of the parameter values. This assumption is valid only in the vicinity of MAP values, unless the model is linear in all parameters. Several studies in the past decade, some of which mentioned below, employed sampling techniques to explore non-Gaussian posterior distributions for parameters in ecosystem models.

Knorr and Kattge (2005), Braswell et al. (2005), and Xu et al. (2006) employed Bayesian frameworks to estimate parameters of terrestrial ecosystem models. These studies employed Metropolis–Hastings Markov chain Monte Carlo
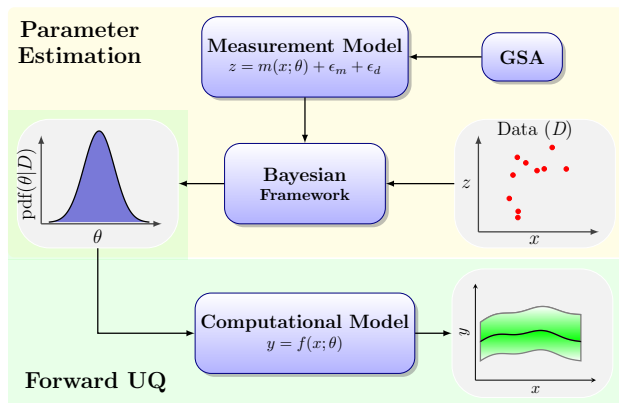
Figure 1. Schematic of parameter estimation, on yellow background, and forward UQ workflows, on green background. For this work DALEC is used as both "measurement model", m, and as "computational model", f. In the Bayesian framework, parameter estimation depends both on the model error ϵ_m and on the measurement error ϵ_d.

(MCMC) techniques to sample the posterior density of model parameters constructed based on eddy covariance measurements of carbon fluxes as well as based on synthetic data sets. Tang and Zhuang (2009) employed both global sensitivity analysis (GSA) and a Bayesian framework to improve parameterization of a terrestrial ecosystem model. This study employed Latin hypercube sampling from the prior density of model parameters, and a sampling-importance resampling method to construct posterior densities for model parameters. Ricciuto et al. (2008) employed an MCMC approach to sample the posterior densities of key parameters for combined global-scale terrestrial and ocean carbon cycle models. The study found that temporal correlation has a significant impact on the calibrated parameters and subsequently on model predictions. A recent review by Zobitz et al. (2011) provides a primer on data assimilation studies with MCMC.

Several studies compared several parameter estimation methods for terrestrial biogeochemical models. Participants in the OptIC (Optimisation InterComparison) project (Trudinger et al., 2007) presented results employing optimization, variational, and sampling methods. Similarly, the REFLEX (REgional FLux Estimation eXperiment) project (Fox et al., 2009) selected the data assimilation linked ecosystem carbon (DALEC) v1 model (Williams et al., 2005) to assess the performance of several parameter estimation algorithms, using both synthetic and observed net ecosystem exchange (NEE) and leaf area index (LAI) data. More recently, Ziehn et al. (2012) compared variational and sampling techniques to estimate parameters for BETHY (Biosphere Energy Transfer and Hydrology), a process-based model of the terrestrial biosphere.

From this review, we noted a set of critical outstanding research questions in the context of constraining carbon cy-

cle models. First, few, if any, calibration studies have investigated steady-state/transient assumptions. It is also important for the ecological community to understand how information content depends on model assumption, e.g., steady state vs transient. Second, carbon cycle models require a complete parameter sensitivity analysis, particularly with respect to temporal dynamics. Such analyses are vital for organizing effective parameter calibration followed by an estimation of the predictive skill of ecosystem models.

In this paper we propose a Bayesian framework for the estimation of uncertainties in ecosystem land model parameters followed by a forward uncertainty quantification (UQ) study to examine the predictive capabilities of the model given the calibrated set of parameters. The Bayesian formulation provides a flexible framework for handling heterogeneous information, and allows for sequential updates of posterior distributions as the prior information is revised.

Figure 1 shows a schematic of this framework, consisting of two intrinsically connected workflows, for *parameter estimation* and *forward UQ*. In this schematic, the data assimilation linked ecosystem carbon (DALEC) model (Williams et al., 2005) is used for both the "measurement model" $m()$ and the "computational model" $f()$. We employ two model setups in our analysis. In the first approach, DALEC is run in a spin-up mode until the carbon pools reach a quasi-steady state. In the second approach, each ecosystem model run consists of one cycle only. In this approach the carbon pools are part of the investigation on model parameters, either for the purpose of estimating densities of model inputs or to propagate these densities forward to model outputs. More details on the steady-state/transient model setups are provided in Sect. 2.

To facilitate the estimation of a high-dimensional posterior density for model parameters, we first rank the importance of specific model parameters on model outputs via GSA. Specifically we employ variance-based decomposition techniques to compute Sobol indices (Sobol, 1993; Campolongo et al., 2000). Posterior densities are estimated first for the most important parameters, while less important parameters are fixed at their nominal values. This constraint is subsequently relaxed to arrive at a joint posterior distribution over the entire parameter space. Finally, we undertake a Bayesian posterior predictive check (Lynch and Western, 2004) to assess the adequacy of the calibrated carbon model to predict the experimental observations. The predictive skill of this model is further assessed via continuous rank predictive score (CRPS) (Gneiting and Raftery, 2007) computations. The analysis steps mentioned here are undertaken with the help of the UQ toolkit (UQTk)[1]. UQTk is a collection of software libraries and tools for the quantification of uncertainty in numerical model predictions. A self-contained package containing specific UQTk libraries and scripts tailored for the
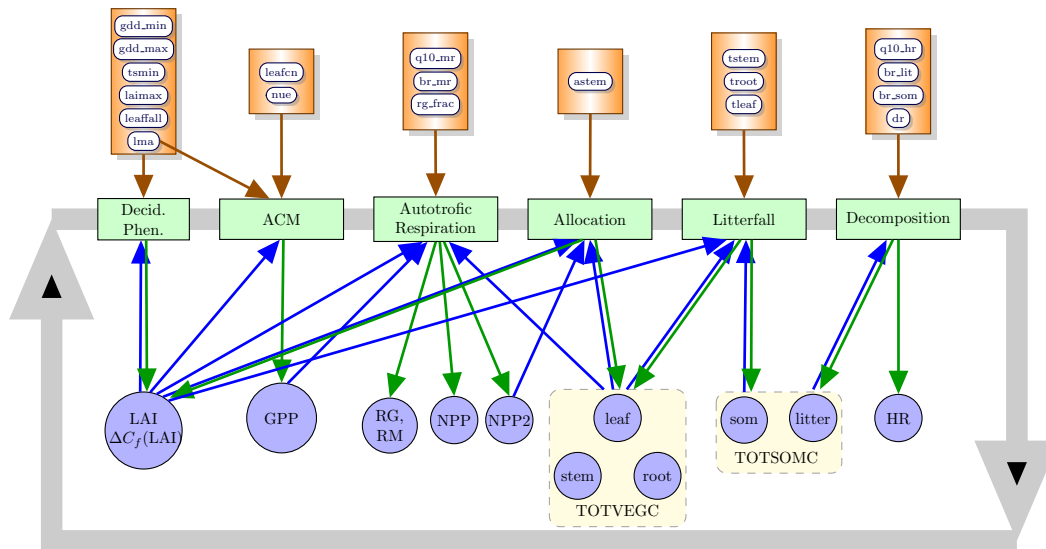
---

[1]http://www.sandia.gov/UQToolkit

Figure 2. Schematic of processes, shown with green boxes, in DALEC with associated parameters, listed in orange boxes. The blue arrows indicate how internal parameters and QoIs, shown with blue circles, impact DALEC processes, while the green arrows show the impact of processes on the QoI and other internal parameters.

workflows presented is this paper is available in the Supplemental material.

This paper is organized as follows. Section 2 provides a description of the processes comprising DALEC and of their associated parameters. Section 3 presents the GSA results, including total-order effects in Sect. 3.1 and joint effects in Sect. 3.2. Posterior densities for model parameters are explored in Sect. 4 and the predictive capabilities are estimated in Sect. 5. We end with conclusions in Sect. 6.

2 Description of the carbon cycle model

The schematic in Fig. 2 shows a 1-day time step consisting of a sequence of process-based sub-models shown with green boxes. These sub-models are connected via fluxes and interact with five major carbon pools. The fluxes calculated on any given day impact carbon pools and processes in subsequent days. The blue arrows in this figure indicate carbon pools or model variables that are input parameters to specific sub-models, while green arrows indicate the carbon pools or model variables affected by a particular sub-process.

The version of DALEC used in this study is based on a modified version of the DALEC v1 model (Williams et al., 2005; Fox et al., 2009). The model has been modified to facilitate comparisons with the community land model (Thornton et al., 2007), and with the Local Terrestrial Ecosystem Carbon Model (Ricciuto et al., 2011)[2]. It consists of three vegetation carbon pools, for leaf, stem, and root, and two soil car-

bon pools, for soil organic matter and litter. Photosynthesis is driven by the aggregate canopy model (ACM) (Williams et al., 2005), which itself is calibrated against the soil–plant–atmosphere (SPA) model (Williams et al., 1996). ACM was updated to employ a temperature-based deciduous phenology used in Ricciuto et al. (2011), driven by the six parameters shown in Fig. 2. Spring phenology is driven by a linear relationship to growing degree days, while senescence is driven by mean air temperature. To reduce model complexity, the plant labile pool was removed and stem carbon is used to support springtime leaf flush given the Spring phenology and the maximum LAI parameter. Given the importance of maintenance respiration in other sensitivity analyses (Sargsyan et al., 2014), this process was added along with parameters controlling the base rate and temperature sensitivity.

In this version of DALEC, ACM shares one parameter, the specific leaf area (*lma*), with the deciduous phenology and employs two additional parameters, leaf C : N ratio (*leafcn*) and nitrogen use efficiency (*nue*). The autotrophic respiration model computes the growth and maintenance respiration components and is controlled by three parameters: the growth respiration fraction (*rg_frac*), the base rate at 25 °C (*br_mr*), and temperature sensitivity for maintenance respiration (*q10_mr*). The allocation sub-model partitions carbon to several vegetation carbon pools. Leaf allocation is first determined by the phenology sub-model, and the remaining available carbon is allocated to the root and stem pools depending on the fractional stem allocation parameter (*astem*). The "litterfall" sub-model redistributes the carbon content from vegetation pools to soil pools and is based on the turnover

---

[2]The source code for the modified DALEC version is available in the Supplemental material.

Table 1. Description of model parameters.

	Param.	Nom. val.	Range	Description	Units
Decid. Phen.	*gdd_min*	100	10–250	threshold for leafout	[°C day]
	gdd_max	200	50–500	threshold for max. LAI	[°C day]
	tsmin	5	0–10	temperature for leaf fall	[°C]
	laimax	4	2–7	seasonal max. LAI	[m^2 leaf/m^2]
	leaf fall	0.1	0.03–0.95	rate of leaf fall	[day^{-1}]
	lma	80	20–150	specific leaf area	[g C/m^2 leaf]
ACM	*leafcn*	25	*fixed*	leaf C : N ratio	[g C/g N]
	nue	7	1–20	nitrogen use efficiency	[ ]
A.R.	*q10_mr*	2	1–4	maintenance resp. T-sensitivity	[ ]
	br_mr	10^{-4}	10^{-5}–10^{-2}	base rate for maintenance resp.	[g C m^{-2} day^{-1}/g C biomass]
	rg_frac	0.2	0.05–0.5	growth respiration fraction	[ ]
A.	astem	0.7	0.1–0.95	allocation to plant stem pool	[ ]
Litter.	*tstem*	$\frac{1}{50 \times 365}$	$\frac{1}{250 \times 365} - \frac{1}{10 \times 365}$	stem turnover time	[day^{-1}]
	troot	$\frac{1}{5 \times 365}$	$\frac{1}{25 \times 365} - \frac{1}{365}$	root turnover time	[day^{-1}]
	tleaf	10^{-2}	10^{-3}–10^{-1}	leaf turnover time	[day^{-1}]
Decomp.	*q10_hr*	2	1–4	heterotrophic resp. T-sensitivity	[ ]
	br_lit	$\frac{1}{2 \times 365}$	$\frac{1}{5 \times 365} - \frac{10}{5 \times 365}$	base turnover for litter	[g C m^{-2} day^{-1}/g C litter]
	br_som	$\frac{1}{30 \times 365}$	$\frac{1}{100 \times 365} - \frac{1}{10 \times 365}$	base turnover for SOM	[g C m^{-2} day^{-1}/g C SOM]
	d*r*	10^{-3}	10^{-4}–10^{-2}	decomposition rate	[day^{-1}]

times for stem (*tstem*), root (*troot*), and leaves (*tleaf*). The sequence of sub-models concludes with the "decomposition" which models the heterotrophic respiration component and the decomposition of litter into soil organic matter (SOM). This sub-model is driven by temperature sensitivity for heterotrophic respiration (*q10_hr*), the base turnover times for litter and SOM at 25 °C (*br_lit*, *br_som*, respectively), and by the decomposition rate (*dr*) from litter to SOM.

Model parameters and their nominal values are provided in Table 1. These parameters are grouped according to the sub-model that employs them. Except for leaf mass per unit area (*lma*) which impacts both the deciduous leaf phenology and ACM, all other parameters are employed in single sub-models. The numerical ranges and nominal values for these parameters are provided in the table, and are designed to reflect average values and broad uncertainties associated with the temperate deciduous forest plant functional type that includes Harvard Forest (Fox et al., 2009; White et al., 2000; Ricciuto et al., 2011). In addition to the model parameters, several processes are driven by the observed air temperature, solar radiation, vapor pressure deficit, and CO_2 concentration at the flux tower site.

As mentioned in the Introduction, for this study we consider two approaches for running DALEC. The first approach employs a steady-state assumption, with DALEC run in a spin-up mode until it reaches a quasi-steady state. For this study we declare the model to be in a quasi-steady state when the relative L_2 error between successive cycles becomes less

than a threshold value of 10^{-6} for select model outputs. For the range of parameters employed in the runs presented here, the model spin-up takes typically 30–50 cycles of the 1992–2006 meteorology (450–750 total years) depending on the parameter values, especially the turnover time of slow carbon pools. In this context, each cycle corresponds to running the model for 15 years with the meteorology inputs of 1992–2006. At the start of the first cycle, the carbon pools are initialized to zero with the exception of stem carbon, which is set at a value to "seed" leaf growth in the following season. For subsequent cycles, the carbon pools are initialized with the final state from the previous cycle. The daily quantities of interest (QoIs) output by DALEC in the first cycle after the system reaches a steady state are used for several analyses presented in this paper. This approach follows the protocol for the North American Carbon Program (NACP) interim synthesis simulations, but fails to capture, for example, the large negative NEE observed at Harvard Forest. In the second approach, the initial values of the carbon pools in January 1992 are added to the set of model parameters to be estimated. This approach employs transient assumptions and, for any given set of parameter values, DALEC is run one cycle only, for 1992–2006. The resulting model output values are then used to study the model behavior under transient conditions. The model evaluations are cheaper compared to the first approach; however, the dimensionality of the parameter space of DALEC is increased by 5, with 3 vegetation carbon pools and 2 soil carbon pools, from 18 to 23 parameters.

Henceforth, we will refer to these two approaches as D_{ST} and D_{TR}.

3 Global sensitivity analysis

GSA formally studies how the change in model output can be apportioned to changes in the model inputs. Given our focus on statistical model calibration and UQ, we employ variance-decomposition methods where the variance of the model output is decomposed into fractions associated with input factors and their interactions. The primary QoI for GSA is NEE, for which we have experimental observations available. We explore GSA for several other QoIs to understand the role each parameter or set of parameters plays in determining other QoIs in addition to NEE. Specifically we consider the gross primary production (GPP), the total vegetation carbon (TVC), and the total soil carbon (TSC).

The effects of input parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_{N_\theta}\}$ and their interactions on a model output $y = m(\boldsymbol{\theta})$, are quantified through Sobol indices (Sobol, 1993; Campolongo et al., 2000). The first-order Sobol indices are given by

$$S_i = \frac{\mathrm{Var}_{\theta_i}[\mathbb{E}_{\boldsymbol{\theta}_{\sim i}}(m(\boldsymbol{\theta})|\theta_i)]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]}, \quad i = 1, \ldots, N_\theta, \tag{1}$$

where $\boldsymbol{\theta}_{\sim i} = \{\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_{N_\theta}\}$, $\mathbb{E}_{\boldsymbol{\theta}_{\sim i}}[\cdot]$ is the expectation with respect to $\boldsymbol{\theta}_{\sim i}$, and $\mathrm{Var}_{\theta_i}[\cdot]$ is the variance with respect to θ_i. Note that, in this context, sub-script i can denote one parameter or a group of parameters. Such a group, corresponding to the phenology model, is presented below.

Similarly, the joint sensitivity indices S_{ij} are

$$S_{ij} = \frac{\mathrm{Var}_{\theta_i,\theta_j}[\mathbb{E}_{\boldsymbol{\theta}_{\sim(i,j)}}(m(\boldsymbol{\theta})|\theta_i,\theta_j)]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]} - S_i - S_j,$$
$$i, j = 1, \ldots, N_\theta. \tag{2}$$

While interactions between three or more parameters can be defined in a similar fashion, for most physical models these higher-order interactions are negligible.

The sensitivity index S_i can be interpreted as the fraction of the variance in the QoI that can be attributed to the ith input parameter only, while S_{ij} is the variance fraction that is due to the joint contribution of the ith and jth input parameters. The total sensitivity index combines the first-order sensitivity indices with joint sensitivity and higher-order interactions to yield

$$S_i^T = S_i + \sum_{\substack{j \\ i \neq j}} S_{ij} + \sum_{\substack{j,k \\ i \neq j \neq k \neq i}} S_{ijk} + \ldots$$
$$= \frac{\mathbb{E}_{\boldsymbol{\theta}_{\sim i}}[\mathrm{Var}_{\theta_i}(m(\boldsymbol{\theta})|\boldsymbol{\theta}_{\sim i})]}{\mathrm{Var}_{\boldsymbol{\theta}}[m(\boldsymbol{\theta})]}. \tag{3}$$

This index measures the fractional contribution to the total variance due to parameter θ_i and all interactions with all other model parameters.

Starting from the derivation of these indices, based on the decomposition of variance, the sum of all first-order indices and joint and higher-order interaction indices sums to one

$$1 = \sum_i S_i + \sum_{\substack{i,j \\ i \neq j}} S_{ij} + \ldots.$$

Given that all Sobol indices are greater or equal to zero, it follows that $\sum_i S_i \leq 1$. The reverse is true for the total effect indices, $\sum_i S_i^T \geq 1$, due to multiple counting of joint and higher order parameter interactions.

Total effect indices are useful to ascertain which parameter or group of parameters has the most impact on a particular QoI, and also decide which parameters are less important and can potentially be fixed at their nominal value without a significant impact on the model output. Joint sensitivity indices can be used to verify or discover interactions between the computational model components as related to a specific model output. In this paper we will present results for total effect and joint sensitivity Sobol indices, while skipping first-order Sobol indices for brevity.

The Sobol indices Eqs. (1)–(3) can be written in integral forms, but these integrals are not analytically tractable when the input parameter space is high dimensional. In order to evaluate these indices numerically we employ a Monte Carlo approach enhanced by techniques described by Saltelli (2002) and modified by Kucherenko et al. (2012) to account for parameter dependencies. This method employs sampling of the input parameters from their prior distributions and an efficient re-use of model evaluations to reduce the computational cost of estimation of the above conditional variances.

We employ informative priors, described in Sect. 4.3, for all model parameters. The prior distributions for all parameters are assumed independent, except for the Spring phenology parameters *gdd_min* and *gdd_max*, which are bound by the inequality constraint *gdd_min* < *gdd_max*. Consequently, for these two parameters we will compute a compound sensitivity index; i.e., S_i^T for $i = (gdd\_min, gdd\_max)$ which is the total effect index based on joint prior distribution of this set of parameters, including all interactions between either *gdd_min* or *gdd_max*, or both, and the rest of the DALEC parameters.

For each of the QoIs mentioned above, we compute monthly averages corresponding to the entire simulation; i.e. the January average is computed using the January daily QoI values for all available years. The simulations are driven by daily minimum and maximum temperatures, global radiation, and CO_2 concentration for years 1992–2006, at the Harvard Forest site (Urbanski et al., 2007).

3.1 Total effect indices

Figures 3–6 show matrices of total effect indices, S_i^T, for the four QoIs mentioned above. Each row in these matrices shows the indices corresponding to a particular monthly average QoI.
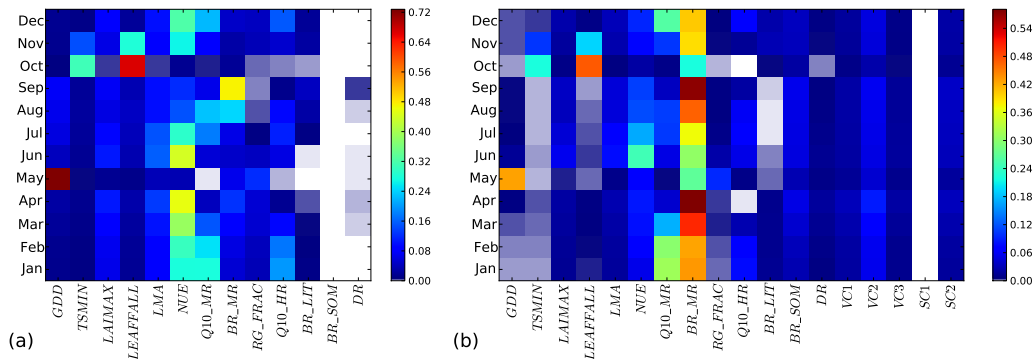
Figure 3. Matrices with total effect Sobol indices, S_i^T, for monthly averages of NEE for **(a)** D_{ST} and **(b)** D_{TR}. The colormap changes from red for large index values to blue for indices $\approx 1\%$. The grayscale corresponds to Sobol index values from 1 % down to 0.1 %, while blank cells indicate values smaller the 0.1 %.
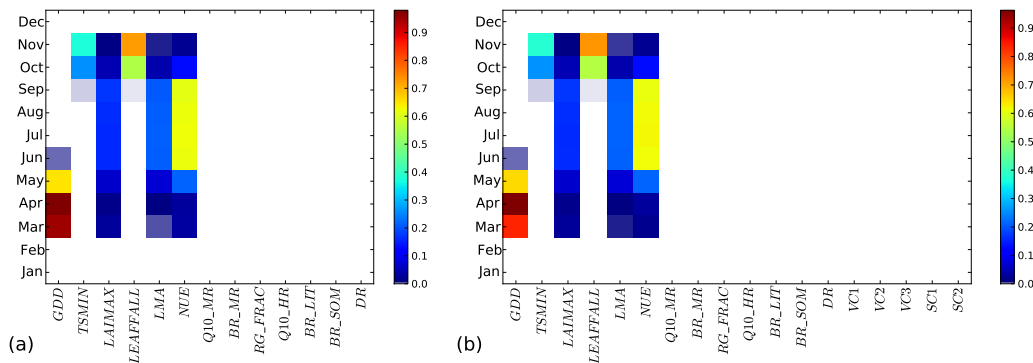

Figure 4. Matrices with total effect Sobol indices, S_i^T, for monthly averages of GPP for **(a)** D_{ST} and **(b)** D_{TR}. The colormap setup is similar to the one in Fig. 3.

Different parameters have larger impacts at certain times of the year. For NEE corresponding to D_{ST}, in Fig. 3a, phenology parameters *tsmin* and *leaf fall*, which control the senescence of leaves in the Fall, have a significant impact on NEE during this period only. Specifically, *tsmin*, which is the critical temperature at which *leaf fall* begins, mainly affects NEE in October. For D_{TR}, in Fig. 3b, the base rate of maintenance respiration *br_mr*, which represents a carbon cost plants must continuously spend during their lifetime, becomes the dominant parameter for NEE. In the transient configuration, the autotrophic respiration sub-model controls most of NEE variance. The total effect index for several parameters, i.e. *astem*, *tstem*, *troot*, and *tleaf*, are not shown in this figure, since they have a negligible contribution to NEE variance.

Similar behavior is seen for parameters that control GPP, in Fig. 4. Parameter *gdd_min*, which is part of the pair *gdd*=(*gdd_min*,*gdd_max*) in this figure, is the number of growing degree days at which leaf budbreak occurs. This parameter has the most impact in March and April. The strong dependence of these fluxes on phenology parameters highlights the importance of an accurate phenology model, as

has been shown in other modeling studies, (e.g., Richardson et al., 2012). On the other hand, the nitrogen use efficiency *nue*, which controls the amount of GPP per unit leaf nitrogen, is important throughout most of the growing season (June–September). This is broadly consistent with other sensitivity studies that have shown strong sensitivity to leaf nitrogen, e.g., Sargsyan et al. (2014). Unlike for NEE, the GPP fluxes exhibit a similar dependence on the parameters controlling the phenology and aggregate canopy modes for both D_{ST} and D_{TR}.

TVC and TSC are carbon pools and tend to vary on a much larger timescale than GPP or NEE, which are fluxes. Therefore, the Sobol indices do not exhibit significant seasonal variability. TVC is a sum of three carbon pools, *vc1* (for leaf C), *vc2* (for stem C), and *vc3* (for root C). For both D_{ST} and D_{TR}, in Fig. 5, this QoI is most strongly controlled by the base rate of maintenance respiration *br_mr*. For D_{TR}, the initial value of *vc2* exhibits a small, but non-negligible, total effect index of about 10 % on the total variance of TVC.

TSC corresponding to D_{ST}, in Fig. 6a, is mostly controlled by both *br_mr* and the base rate of decomposition for soil organic matter *br_som*, which effectively determines the pool
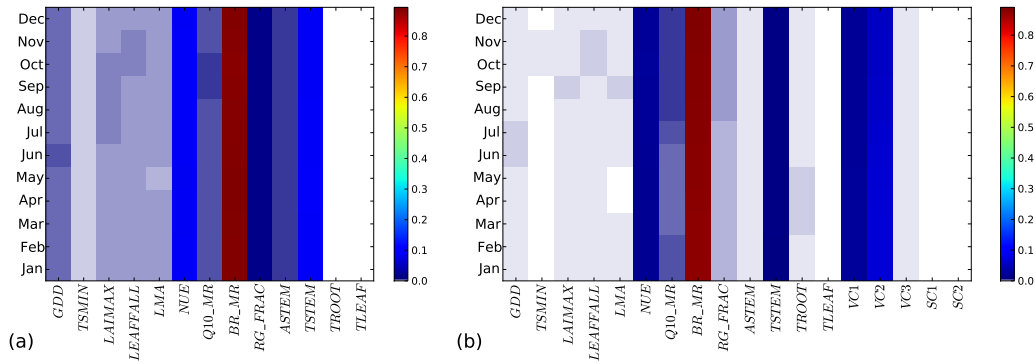
Figure 5. Matrices with total effect Sobol indices, S_i^T, for monthly averages of TVC for **(a)** D_{ST} and **(b)** D_{TR}. The colormap setup is similar to the one in Fig. 3.
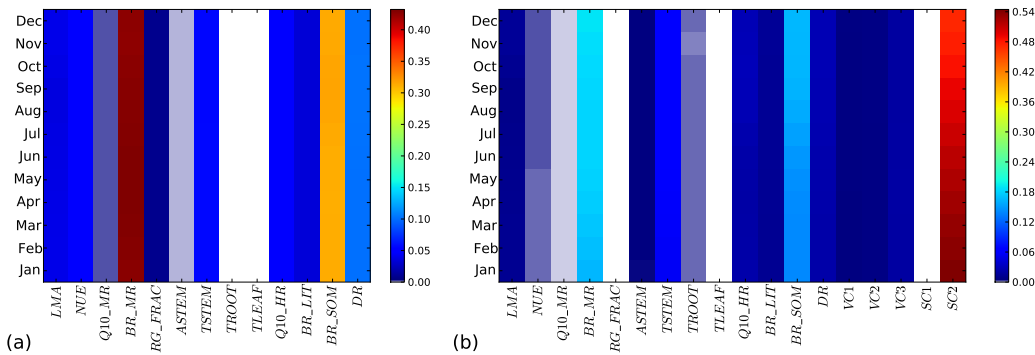

Figure 6. Matrices with total effect Sobol indices, S_i^T, for monthly averages of TSC for **(a)** D_{ST} and **(b)** D_{TR}. The colormap setup is similar to results in Fig. 3.

residence time. Given the same inputs, a pool with a longer residence time will contain more carbon. For D_{TR}, in Fig. 6b, the initial value of soil organic matter pool (*sc2*) becomes dominant and exhibits a total effect index of about 50 %. For this setup, the impact of *br_mr* and *br_som* on the total variance of TSC is about 40 %, down from about 80 % for the quasi-steady-state setup for D_{ST}.

The total sensitivity index results indicate that, for some quantities of interest like GPP and TVC, the simulation setup, i.e. D_{ST} vs. D_{TR}, does not change significantly the effect of model parameters on the model outputs. For these two model outputs the dominant parameters are similar for both setups, given the priors employed for the model parameters, including the carbon pools for D_{TR}. Unlike for GPP and TVC, the simulation setup changes the relative importance of model parameters on NEE and TSC. This takes place either through a change in the relative importance of phenology and ACM model parameters (for NEE) or by bringing a significant contribution from the carbon pools (for TSC). In the next section we examine joint effect indices for parameter pairs to determine what fraction of the total effect indices is due to interactions between model parameters.

3.2 Joint effects

Figures 7–9 show relevant joint sensitivity indices corresponding to the four QoIs examined in this study. In these figures, each node shows relevant parameters while the label on each link corresponds to the joint Sobol index S_{ij}, in % units. The joint sensitivity Sobol index values are rounded to the nearest integer for clarity.

Both NEE and GPP exhibit seasonal variability for the total effect Sobol indices. For these parameters the joint parameter interactions are only relevant during Fall, accounting for about 10–15 % of the total variance in the corresponding QoI, and play an important role in determining the evolution of the carbon cycle during the senescence period. Figures 7 and 8, showing these interactions during October, are representative of results throughout Fall. For both NEE and GPP, the interaction *tsmin* and *leaf fall* is significant during Fall, while interactions between other phenology, ACM, and AR parameters are negligible. In general joint sensitivity maps for NEE and GPP are similar between D_{ST} and D_{TR}.

Similar to the total effect index results for TVC and TSV, the joint sensitivity indices display little seasonal variability. The results shown in Fig. 9 for these QoIs correspond to September and are representative of all monthly averages
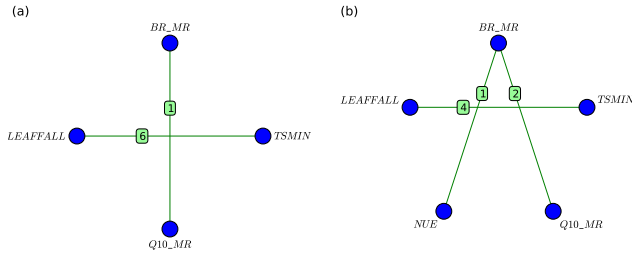
Figure 7. Relevant joint Sobol indices, S_{ij}, corresponding to October NEE averages for **(a)** D_{ST} and **(b)** D_{TR}. The labels on each line show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.

Figure 8. Relevant joint Sobol indices, S_{ij}, corresponding to October GPP averages for **(a)** D_{ST} and **(b)** D_{TR}. The labels on each line show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.
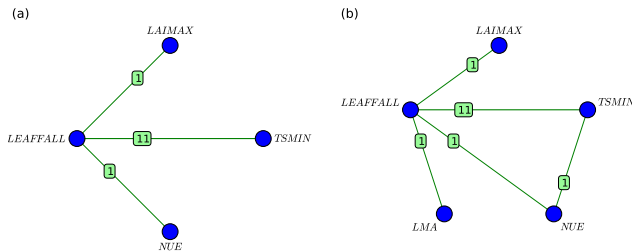
(results not shown). Moreover, only D_{ST} results are shown in this figure since the corresponding D_{TR} results are almost identical to D_{ST}. For TVC the data in Fig. 9a indicates that the interaction between AR (through *br_mr*) and ACM (*nue*) and litterfall (*tstem*) sub-models contribute about 10 % to TVC variance. In fact these joint interactions represent about half of the total effect index of *nue* and *tstem*, shown in Fig. 5. The results in Fig. 9b show that the interactions between model parameters are important for TSC as well. For this QoI, the interaction AR (*br_mr*) and decomposition (*br_som*) sub-models account for about 10–30 % of the corresponding total effect index values, shown in Fig. 6.

The GSA results can be used to understand the effect of model parameters on particular QoIs and discard, from the analysis, parameters that have a negligible impact. In this study, we will use the GSA results to facilitate the calibration of model parameters, by grouping parameters into subsets according to their effect on the relevant QoIs. More details are presented in the following section.

4 Parameter calibration

We employ a Bayesian framework to compute posterior probabilities for model parameters discussed in the previous sections. This framework is well-suited for dealing with uncertainties from different sources, including parametric
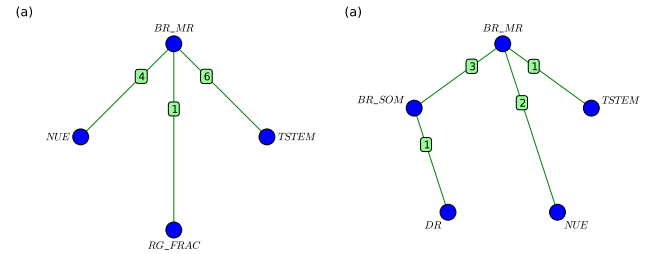

Figure 9. Relevant joint Sobol indices, S_{ij}, corresponding to September averages for **(a)** TVC and **(b)** TSC. Both sets of results are based on D_{ST}. The labels on each line show the magnitude, in %, of Sobol indices for the corresponding parameter pairs.

and model uncertainty as well as experimental errors (Sivia, 1996). Bayes' rule is given as

$$p(\boldsymbol{\theta}|\mathcal{D}) = L_{\mathcal{D}}(\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathcal{D}), \tag{4}$$

where $p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta}|\mathcal{D})$ are the prior and posterior probability densities, respectively, for model parameters $\boldsymbol{\theta}$. These densities represent our knowledge of $\boldsymbol{\theta}$ before and after learning from the data $\mathcal{D}$. The likelihood function $L_{\mathcal{D}}(\boldsymbol{\theta}) = p(\mathcal{D}|\boldsymbol{\theta})$ is the likelihood of the data $\mathcal{D}$ for a particular instance of model parameters $\boldsymbol{\theta}$. The denominator in Eq. (4), $p(\mathcal{D})$, is the "evidence", computed by integrating the numerator over the support of $p(\boldsymbol{\theta})$. It plays the role of a normalizing constant in the parameter estimation context, and is not computed here.

4.1 Calibration data

The data available for the calibration of model parameters consist of the Harvard Forest's daily NEE values processed for the NACP site synthesis study (Barr et al., 2013) based on flux data measured at the site (Urbanski et al., 2007). Hill et al. (2012) estimated that daily NEE estimates follow a normal distribution. The daily observations cover a period of 15 years starting with year 1992. A snapshot of these observations, including the magnitude of the observation error, is provided in Fig. 10. The standard deviations for the daily NEE values were estimated using a bootstrapping technique using half-hourly NEE data (Papale et al., 2006; Barr et al., 2009). The mean standard deviation is about 0.7, with a range of variation between 0.2 and 2.5.

4.2 Likelihood construction

In general, the discrepancy between model predictions and the data can be formalized as

$$z = m(t; \boldsymbol{\theta}) + \epsilon_m + \epsilon_d. \tag{5}$$

Here, t is the time in day units and z is the daily NEE observation described above. Further, ϵ_m is the discrepancy between the model prediction $m(t; \boldsymbol{\theta})$ and the physical truth,
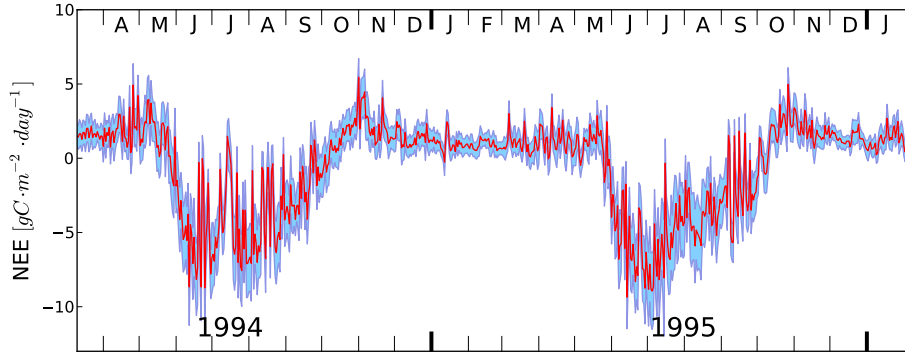
Figure 10. Snapshot of NEE observations (with red line) for the Harvard Forest site. The light-blue region, bordered by thick blue lines corresponds to $\pm 2\sigma$ around the daily NEE values.

while ϵ_d denotes the experimental error. In general it is not straightforward to disambiguate between these two sources of error. For the present study, we presume the experimental error to be known (Papale et al., 2006; Barr et al., 2009). Given that measurements are taken at different times, we further assume that daily measurement noise/errors, ϵ_d, are independent; hence,

$$\epsilon_d \propto N(0, \Sigma_d), \quad \Sigma_d = \mathrm{diag}[\underbrace{\sigma_{d,1}^2, \sigma_{d,2}^2, \ldots}_{N_d}], \quad (6)$$

where N_d is the number of days. Next we will focus our attention on modeling ϵ_m. We propose a multivariate Gaussian distribution, employing a constant bias $\boldsymbol{\mu} = [\mu, \mu, \ldots, \mu]^T$ and a $N_d \times N_d$ square exponential covariance matrix Σ_m with

$$\Sigma_{mi,j} = \sigma_m^2 \exp\left(-(t_i - t_j)^2 / l_c^2\right). \quad (7)$$

Given that t_i is simply a notation for day no. i, the covariance matrix entries are given by $\Sigma_{mi,j} = \sigma_m^2 \exp\left(-(i - j)^2 / l_c^2\right)$, where l_c is a correlation length. This analytical expression for Σ_m is adopted based on the intuition that model errors for successive days are highly correlated while model errors for days that are far apart are uncorrelated. The magnitude of l_c controls the rate of decrease of daily model error correlations.

Given the above formulations of model and data errors, one can group these two into one multivariate normal error term

$$\epsilon = \epsilon_m + \epsilon_d \propto N(\boldsymbol{\mu}, \Sigma), \quad \Sigma_{i,i} = \sigma_m^2 + \sigma_{d,i}^2,$$
$$\Sigma_{i,i\pm k} = \sigma_m^2 \exp\left(-k^2 / l_c^2\right) \quad (8)$$

and the likelihood $L_\mathcal{D}(\boldsymbol{\theta})$ is written as

$$L_\mathcal{D}(\boldsymbol{\theta}) = (2\pi)^{-N_d/2} |\Sigma|^{-1/2}$$
$$\exp\left(-(z - m - \mu)^T \Sigma^{-1} (z - m - \mu)\right). \quad (9)$$

Here, $z = [z_1, z_2, \ldots]$ is the vector of NEE observations, $m = [m(t_1, \boldsymbol{\theta}), m(t_2, \boldsymbol{\theta}), \ldots]$ is the vector of model NEE predictions, and $\boldsymbol{\mu}$ is the model bias vector described above. All

these vectors are N_d long. In addition to the model parameters $\boldsymbol{\theta}$, we now have three additional hyperparameters characterizing the model error: the model bias μ, model error standard deviation σ_m, and correlation length l_c. Unlike for DALEC parameters, for which we employ informed priors described in the next section, for these hyperparameters we employ uninformed priors.

In practice, estimating the likelihood $L_\mathcal{D}(\boldsymbol{\theta})$ can be costly, and prone to numerical instabilities when considering the full $N_d \times N_d$ covariance matrix Σ. Therefore, we will work with band-diagonal covariance matrices, obtained by setting the diagonals of the model error covariance matrix Σ_m to zero beyond a certain bandwidth k_b

$$\Sigma_{i,i\pm k} = 0 \text{ for } k > k_b. \quad (10)$$

The effect of covariance matrix bandwidth on the model error terms $\{\mu, \sigma_m, l_c\}$ and DALEC parameters is studied in Sect. 4.4.1.

4.3 Parameter priors

Following LeBauer et al. (2012), we proceed to construct informed priors for the DALEC model parameters as well as for the initial carbon pool amounts employed in D_{TR}. Considering the nominal values and bounds presented in Table 1, we separate model parameters into two categories. In the first category we place parameters with a range that spans approximately 1 order of magnitude or less. For these parameters we employ truncated normal densities as priors, with the mode set at the nominal values and standard deviations set to one-eighth of the range of variation for each parameter.

In the second category we place parameters for which the range of variation spans more than 2 orders of magnitude. For these parameters we set truncated lognormal density priors. Similarly to the first set of parameters, the parameters of these densities are set such that the mode occurs at the nominal value and the standard deviation is set to one-eighth of the range of variation for each corresponding parameter in

Table 2. Prior setup for the initial carbon pool amounts employed in D_{TR}.

Name	ID	Mean	SD	Constraints
leaf C	$vc1$	0	20	$0 < vc1$
stem C	$vc2$	9000	1800	$0 < vc2$
root C	$vc3$	1500	300	$0 < vc3$
litter C	$sc1$	10	25	$0 < sc1 < 1000$
som C	$sc2$	8800	1760	$0 < sc2$

this category. For both normal and lognormal densities, we truncate the priors based on the ranges presented in Table 1.

For all parameters, except the pair *gdd_min*, *gdd_max*, we consider independent prior distributions. For the growing degree days parameters, given the inequality constraint *gdd_min* < *gdd_max*, we employ a truncated joint normal density setup as a product of one-dimensional normal densities for both *gdd_min* and *gdd_max*. This joint density is appropriately scaled so that it integrates to one over non-rectangular space (due to the inequality constraint) for these two parameters. Similarly, the truncated normal and lognormal densities for the other model parameters are appropriately scaled to account for the finite parameter ranges.

For D_{TR}, the initial carbon pool amounts (representing values on 01 January 1992) are also estimated in addition to the DALEC parameters and the hyperparameters defining the model error. For the carbon pool initial values we also employ truncated normal and lognormal densities. These prior distributions are informed by site observations (Table 2). The initial leaf carbon (*vc1*) is set to zero with a small standard deviation because of the starting date of the simulation, which is in mid-winter well after leaf fall. Initial litter and soil organic mean (*sc1*, *sc2*) values and standard deviations are taken from Gaudinski et al. (2000), while stem carbon is estimated from Urbanski et al. (2007). Specifically, we employ truncated normal densities for all carbon pools except litter carbon (*sc1*). For *sc1*, the mean and the range differ by 2 orders of magnitude; hence, we employ a truncated lognormal density for this pool.

4.4 Posterior distributions via MCMC

A MCMC algorithm is used to sample from the posterior probability density $p(\boldsymbol{\theta}|\mathcal{D})$ in Eq. (4). MCMC is a class of techniques that allows for sampling from a probability density by constructing a Markov chain that has the target density as its stationary distribution (Gamerman, 1997; Gilks et al., 1996). In particular, we employ an adaptive Metropolis algorithm (Haario et al., 2001), which uses the covariance of the previously visited chain states to find better proposal distributions, allowing it to explore the posterior distribution in an efficient manner. Haario et al. (2001) show that, for Gaussian distributions, the adaptive sampling algorithm is similar in performance to the Metropolis algorithm. For
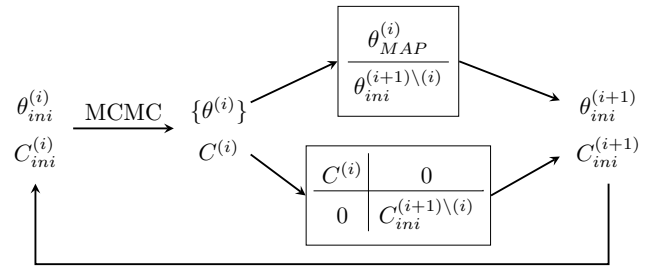

Figure 11. Schematic of the iterative process for parameter calibration. The MCMC sampling of the joint density for the set of parameters $\theta^{(i)}$ starts at $\theta_{\mathrm{ini}}^{(i)}$ using an initial proposal covariance $C_{\mathrm{ini}}^{(i)}$. For the following iteration, $(i+1)$, the initial condition is constructed using the MAP estimate for $\theta^{(i)}$, augmented with initial conditions, in this case the nominal values, for the rest of parameters, $\theta^{(i+1)\smallsetminus(i)}$. The initial proposal covariance $C_{\mathrm{ini}}^{(i+1)}$ is constructed based on the sample covariance matrix for $\theta^{(i)}$, augmented with an initial proposal covariance for $\theta^{(i+1)\smallsetminus(i)}$, $C_{\mathrm{ini}}^{(i+1)\smallsetminus(i)}$.

non-Gaussian posterior densities, the adaptive procedure is superior to non-adaptive procedures; however, the adaptive procedure is challenged by the dimensionality of the parameter space.

To facilitate the convergence of the adaptive MCMC algorithm we proceed gradually, starting with a group of parameters identified as important for NEE through GSA in Sect. 3. The schematic in Fig. 11 shows one iteration in the sequence of MCMC simulations. We also add the model error hyperparameters, in addition to select DALEC parameters, to start the first iteration:

$$\theta^{(1)} = \{gdd\_min, gdd\_max, tsmin, leaffall, nue, q10\_mr, br\_mr\}$$
$$+ \{\mu, \sigma_{\mathrm{m}}, l_{\mathrm{c}}\}$$

with initial values $\theta_{\mathrm{ini}}^{(1)}$ set to the nominal conditions provided in Table 1 for DALEC parameters, and $\mu = 0$, $\sigma_{\mathrm{m}} = l_{\mathrm{c}} = 1$ for model error hyperparameters. The rest of parameters are held constant at their nominal values. The initial covariance matrix, $C_{\mathrm{ini}}^{(1)}$, allows the MCMC algorithm to explore a number of possible states before adapting the sample covariance based on the sample history. For this study we found that a diagonal covariance matrix with entries set to a fraction of about 1/16 of the variances for the corresponding prior density provided a good start for the MCMC algorithm.

The MCMC states obtained during the first iteration are used to compute the covariance matrix corresponding to the first set of parameters $C^{(1)}$ which is then used to construct the initial covariance matrix for the second iteration, $C_{\mathrm{ini}}^{(2)}$. This process is shown schematically in Fig. 11. The initial parameter values for the second iteration consist of the MAP for $\theta^{(1)}$ augmented with the nominal values for

$$\theta^{(2)\smallsetminus(1)} = \{lma, rg\_frac, q10\_hr, br\_lit\}.$$

Table 3. Distance correlation factors for D_{ST}. The diagonal blocks are marked according to the process the parameters contribute to; see also Fig. 2 and Table 1. The entries in the diagonal block show dependencies between parameters from the same process, while the entries in the off-diagonal block show dependencies between parameters from different processes.

	gdd_min	gdd_max	tsmin	laimax	leaf fall	lma	nue	q10_mr	br_mr	rg_frac	astem	tstem	troot	tleaf	q10_hr	br_lit	br_som	dr
gdd_min	1																	
gdd_max	0.6	1																
tsmin	0	0	1															
laimax	0	0.1	0	1														
leaf fall	0.1	0	0.3	0	1													
lma	0	0.2	0	0.1	0.1	1												
nue	0.1	0.3	0	0.1	0.2	0.9	1											
q10_mr	0.1	0.2	0	0	0.1	0.5	0.5	1										
br_mr	0.1	0.2	0.1	0	0.4	0	0	0.1	1									
rg_frac	0.2	0.1	0	0	0.2	0.4	0.7	0.3	0.1	1								
astem	0	0	0	0	0	0	0	0	0	0	1							
tstem	0	0	0	0	0.2	0	0	0.1	0.1	0.1	0	1						
troot	0	0	0	0	0	0	0	0	0	0	0	0.1	1					
tleaf	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
q10_hr	0.1	0.1	0.1	0	0.2	0.2	0.1	0.2	0.2	0.2	0	0.2	0	0	1			
br_lit	0.1	0.1	0.1	0	0.4	0	0.1	0.1	0.5	0.1	0	0	0	0	0.3	1		
br_som	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
dr	0	0	0	0	0.1	0.1	0.1	0.1	0.2	0.1	0	0	0	0	0.1	0.2	0.1	1

The iterative process is completed after the third iteration, with $\theta^{(3)\searrow(2)}$ containing the rest of the DALEC parameters. This iterative algorithm breaks the original high-dimensional problem into a sequence of steps of increasing dimensionality, with each intermediate step starting with a better proposal covariance compared to an approach for which this covariance is empirically chosen.

We employ the Raftery–Lewis diagnostic (Raftery and Lewis, 1992) to determine when the MCMC samples converge to stationary posterior distributions. For D_{ST}, approximately 4×10^6 samples are necessary to predict the 5, 50, and 95 % quantiles of all parameters to within ± 1 % accuracy with 95 % probability. For D_{TR}, the Raftery–Lewis diagnostic test shows that 6×10^6 are necessary for converged posterior densities. Given 5×10^6 MCMC samples, the effective sample size (Kass et al., 1998) (ESS) for D_{ST} varies between 10 000 and 15 000 samples depending on each parameter, while for D_{TR}, ESS is between 8000 and 12 000. This shows that there is significant autocorrelation between chain samples, which is somewhat typical for MCMC samplers in high-dimensional spaces. To ensure converged posterior densities, and since the computational model is cheap, results presented below are based on 7.5×10^6 MCMC samples for both D_{ST} and D_{TR}. When processing the MCMC

samples, we skip the first 10^6 samples, and then "thin" the rest of the samples by picking every 10th sample.

4.4.1 Effect of covariance bandwidth on posterior distributions

We performed several MCMC runs to examine the effect of covariance bandwidth on the estimates of model parameters and hyperparameters. The bandwidth is parameterized by k_b, in Eq. (10), which denotes the number of non-zero diagonals on either side of the main diagonal.

Figure 12a–c show the estimated MAP values for the hyperparameters μ, σ_m, and l_c, corresponding to the model error. In addition to D_{ST} and D_{TR}, we also show results for "D_{TR}^{up}". This run is similar to D_{TR}, except uniform priors with the same range were employed for all carbon pools. The error bars shown in this figure correspond to 2 standard deviations estimated from the MCMC samples.

It seems that the model bias μ, in Fig. 12a, is not significantly affected by the band-diagonal trim of the covariance matrix. For all runs considered here, μ is consistently negative signaling that, on average, DALEC overpredicts the NEE data. The other two model parameters, σ_m and l_c (in Fig. 12b, c), are sensitive to the bandwidth setup, until they reach statistically converged values around $k_b = 10$. The model error standard deviation σ_m mean values for both D_{ST} and D_{TR} are
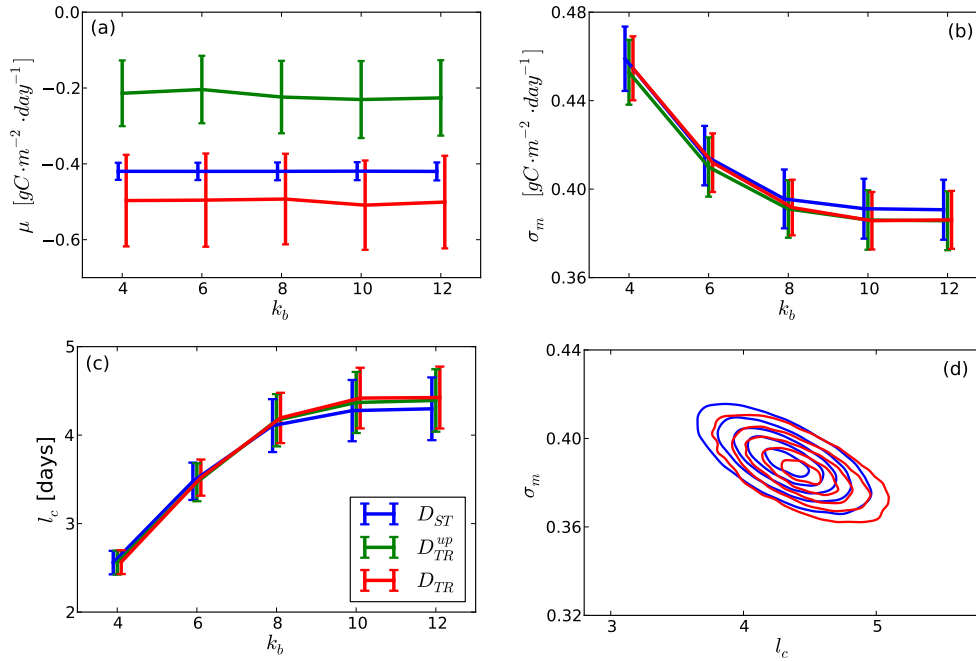
Figure 12. Convergence of model error components with increasing bandwidth of the covariance matrix: **(a)** μ, **(b)** σ_{m}, and **(c)** l_{c}. The joint 2-D marginal density of $(\sigma_{\mathrm{m}}, l_{\mathrm{c}})$ for $k_{\mathrm{b}} = 12$ is shown in **(d)**. In addition to D_{ST} and D_{TR} setups, we also considered "$D_{\mathrm{TR}}^{\mathrm{up}}$", a setup equivalent to D_{TR}, but with uniform priors assumed for the vegetation and soil carbon pools.

slightly below 0.4, compared to a mean value of 0.7 for the NEE measurement error (in Sect. 4.1).

The 2-D joint marginal density for σ_{m} and l_{c}, shown in Fig. 12d for $k_{\mathrm{b}} = 12$, indicate a relatively strong dependence and a negative correlation between these two hyperparameters. Results for larger covariance bandwidths (not shown) confirm that densities of both σ_{m} and l_{c} exhibit converged moments for $k_{\mathrm{b}} > 10$.

It is interesting to note the value for the converged mean correlation length l_{c}. It seems that this hyperparameter does not depend on a particular model setup, at least for the site and time range considered in this study. Further tests, with uniform priors for all parameters, lead to similar mean values for l_{c}. A value of $l_{\mathrm{c}} = 4$, indicating that the model error discrepancy exhibits a timescale of about 8 days, seems to be an intrinsic property of the model. This most likely suggests that model errors follow the variability of NEE over synoptic timescales associated with the periodic passage of weather systems and precipitation events (Mahecha et al., 2010). Further tests, with alternate model error terms, are necessary to verify this observation.

4.4.2 Comparison between D_{ST} and D_{TR}

We first proceed to analyze the model calibration results for D_{ST}, when DALEC is run to a quasi-steady state for each parameter sample. In order to measure the degree of dependence in the posterior distributions for the 18 model parameters, we examine the "distance correlation" values (Székely

et al., 2007) estimated based on the MCMC samples. The distance correlation is a measure of *dependence* between two random variables, being zero when they are independent. Given random variables X and Y with finite first moments, the distance correlation $\mathcal{R}(X, Y) \in [0, 1]$ is defined as

$$\mathcal{R}(X, Y) = \frac{\vartheta^2(X, Y)}{\sqrt{\vartheta^2(X)\vartheta^2(Y)}}, \tag{11}$$

where $\vartheta^2(X, Y)$ is the "distance covariance" between X and Y and $\vartheta^2(X)$ is the "distance variance", $\vartheta^2(X) = \vartheta^2(X, X)$. The distance covariance $\vartheta^2(X, Y)$ is defined as

$$
\begin{aligned}
\vartheta^2(X, Y) = {} & E(\|X - X'\|\|Y - Y'\|) \\
& + E(\|X - X'\|)E(\|Y - Y'\|) \\
& - 2E(\|X - X'\|\|Y - Y''\|),
\end{aligned} \tag{12}
$$

where (X', Y') and (X'', Y'') are independent and identically distributed random variables, with the same joint density as (X, Y). Székely et al. (2007) provide numerical algorithms to compute $\mathcal{R}(X, Y)$ given samples of random variables X and Y. The results are shown in Table 3. In this table, parameters are grouped in blocks according to the sub-model they participate in. The entries in the diagonal blocks show dependencies between parameters in the same sub-model while the entries in off-diagonal blocks indicate dependencies between parameters from different sub-models.

The most important statistical dependencies are between *nue* and *lma* that control the gross photosynthesis (ACM) and
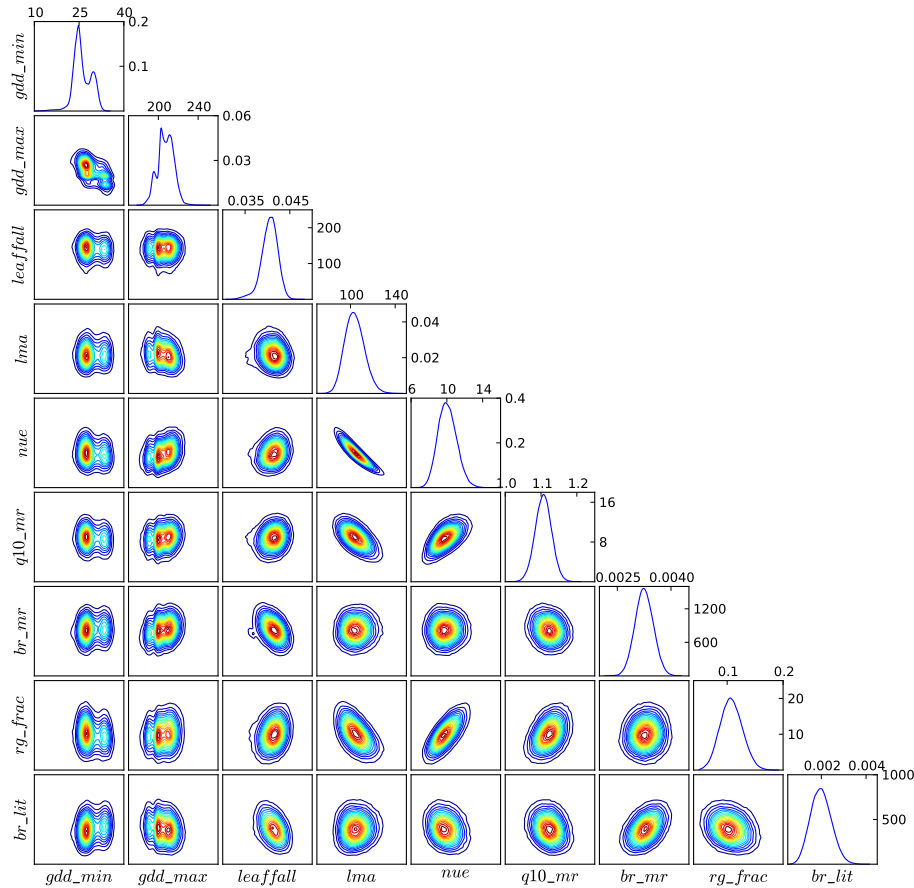
Figure 13. D_{ST} problem: 1-D marginal and 2-D joint marginal probability density functions (PDFs) for parameters showing distance correlation factors above 0.4; see also Table 3. Marginal PDFs are estimated via KDE based on approximately 4×10^5 MCMC samples.

between *rg_frac* and *nue* that control net photosynthesis. Relevant dependencies are also observed between *q10_mr*, a parameter of the autotrophic respiration process, and the gross photosynthesis parameters. In order to further understand the dependencies between model parameters, we compute 1-D and 2-D joint marginal densities, via kernel density estimation (KDE) (Scott, 1992; Silverman, 1986), for the model parameters that exhibit at least one distance correlation factor that is greater than 0.4. These results are shown in Fig. 13. The statistical dependencies identified above through $\mathcal{R}$ are also evident in 2-D joint marginal densities for the same parameters.

Figure 14 shows 1-D marginal densities for the rest of the parameters. These parameters show little dependence on other parameters and so the 1-D marginal distribution is sufficient to characterize their density. Some parameters, e.g., *astem*, *tleaf*, or *br_mr*, show little update from prior to posterior densities. For *br_som*, its turnover rate is slow enough such that the NEE data contain little useful information. For *tleaf*, the lack of information is due to the fact that the effects of leaf turnover on net fluxes are much more strongly controlled by their timing, as determined by the phe-

nology parameters, than by the background turnover rate. The posterior densities for other parameters, e.g., *laimax*, are tilted toward one end of their prior range. This might indicate that the model error term is not sufficient to describe the discrepancy between the model and the data, and the calibration process attempts to compensate for structural discrepancies between observations and model predictions by pushing some parameters toward either the minimum or the maximum value of their prior range.

The posterior density for *tsmin* exhibits an interesting piecewise quasi-linear profile. This is due to the fact that minimum daily temperatures, in degrees Celsius, are provided with 1 decimal digit accuracy and this parameter is a threshold for leaf drop; i.e., its participation in the computational model is through an "if" statement. Hence, all samples between successive one-digit accurate thresholds are equally likely during the MCMC sampling process, and the product between piecewise uniform likelihood and the normal prior results in the posterior density profile observed in Fig. 14.

Next, we analyze the calibration results for D_{TR}. For this model setup, the initial values for the carbon pools at the beginning of year 1992 are part of the set of model parameters
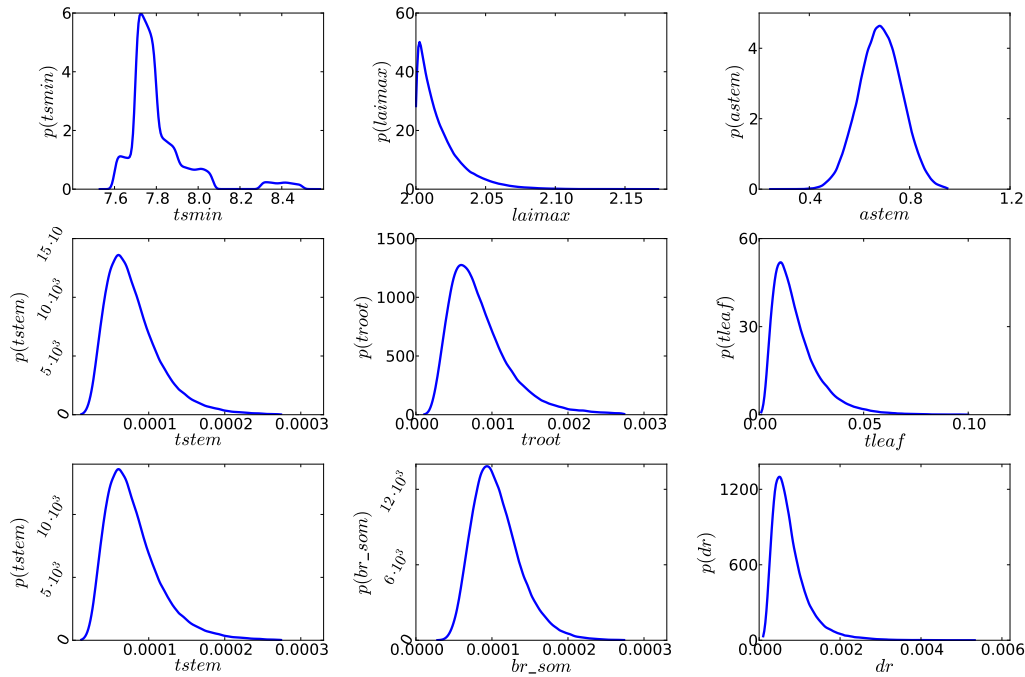
Figure 14. D_{ST} problem: 1-D marginal PDFs for parameters showing distance correlation factors less than 0.4 with other parameters; see also Table 3. Marginal PDFs are estimated via KDE based on approximately 4×10^5 MCMC samples.

and each DALEC simulation consists of only one cycle, for the time span 1992–2006. The distance correlation matrix for D_{TR} parameters that are common to D_{ST} has entries that are by-and-large similar to the ones shown in Table 3 indicating that the dependence between model parameters is not altered by the model setup. This observation is confirmed by visual inspection of the 1-D and 2-D joint marginal densities based on D_{TR} results for the same parameters as the ones shown in Fig. 13 (results not shown).

Finally, Fig. 15 shows marginal densities for two carbon pools that were updated in the calibration exercise D_{TR}. *vc3* corresponds to the stem carbon while *sc1* and *sc2* correspond to the litter carbon and soil organic matter, respectively. Both *vc3* and *sc2* exhibit some dependence on the temperature sensitivities for maintenance respiration and heterotrophic respiration, *q10_mr* and *q10_hr*, respectively. These dependencies are consistent with the flow of information depicted in Fig. 1.

Next we examine the departure of each parameter's posterior density from its prior density as a result of the Bayesian update via Eq. (4). We quantify these changes via the Kullback–Leibler (KL) divergence between prior and marginal posterior densities,

$$D_{KL}(p||q) = \int_{-\infty}^{\infty} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx, \qquad (13)$$

where p is the posterior density and q is the prior density. KL divergence results are presented in Fig. 16. In this figure, parameters are sorted in ascending order based on the D_{KL} val-

ues for D_{ST}. Parameters that exhibit $D_{KL} < 0.5$ for both D_{ST} and D_{TR} are excluded from this figure for clarity. Moreover, the C pools shown in this figure are only present for D_{TR}; hence, there is no D_{ST} result for these parameters. The right half of this figure contains parameters that were identified as important for NEE in Sect. 3. These parameters are well constrained by the NEE data, reflecting the useful information in the flux data, for example on the timing of phenological events (*gdd_min*) and the dynamics of autotrophic respiration (*br_mr*, *q10_mr*). In general D_{KL} results are similar for D_{ST} and D_{TR}, perhaps with the exception of *br_som*. For D_{ST}, the NEE data contain little information on the turnover rate of SOM. For D_{TR}, the inclusion of carbon pools, in particular the SOM pool (*sc2*), impacts the Bayesian update of this parameters due to the dependencies observed in the joint marginal densities, shown in Fig. 15.

5 Predictive assessment

In this section we explore the predictive skill given the posterior distributions for the model parameters for D_{ST} and D_{TR}. First, we employ the Bayesian posterior predictive distribution (Lynch and Western, 2004) to assess the adequacy of the calibrated DALEC model, and the Gaussian data noise model, for prediction of the NEE observations. Specifically, the posterior distribution for the *predicted* NEE data, $p(y|\mathcal{D})$, is computed by marginalization of the likelihood over the posterior distribution of model parameters and
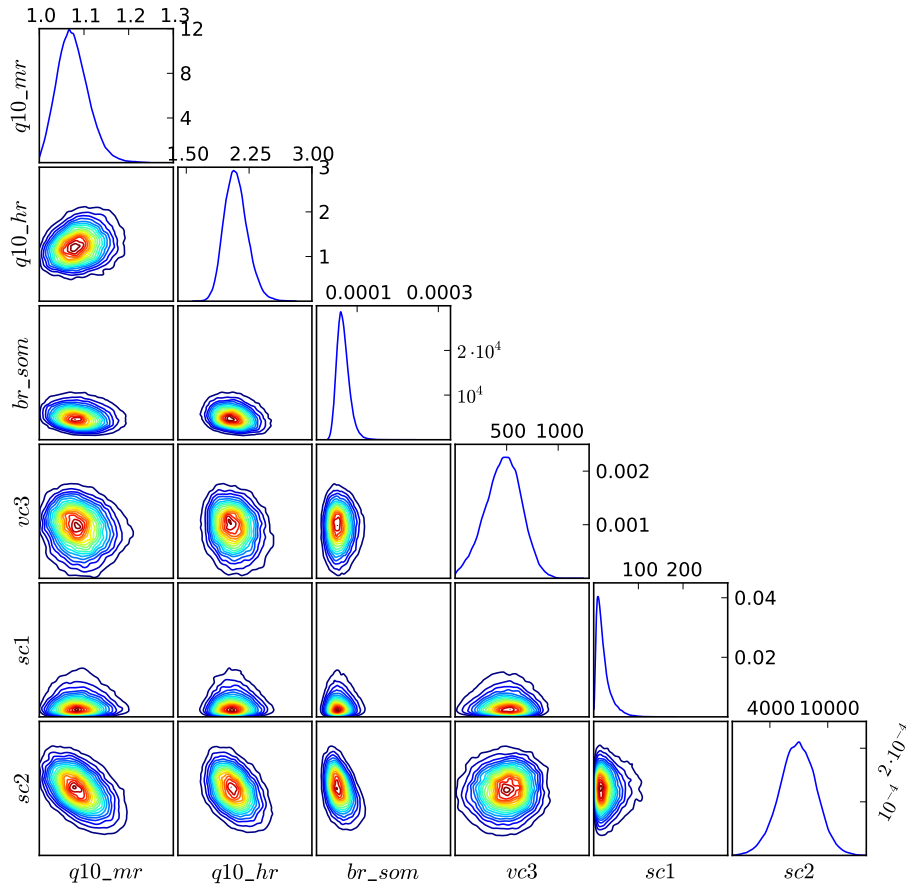
Figure 15. D_{TR} problem: 1-D marginal and 2-D joint marginal PDFs for select parameters correlated with the carbon pools. Marginal PDFs are estimated via KDE based on approximately 4×10^5 MCMC samples.
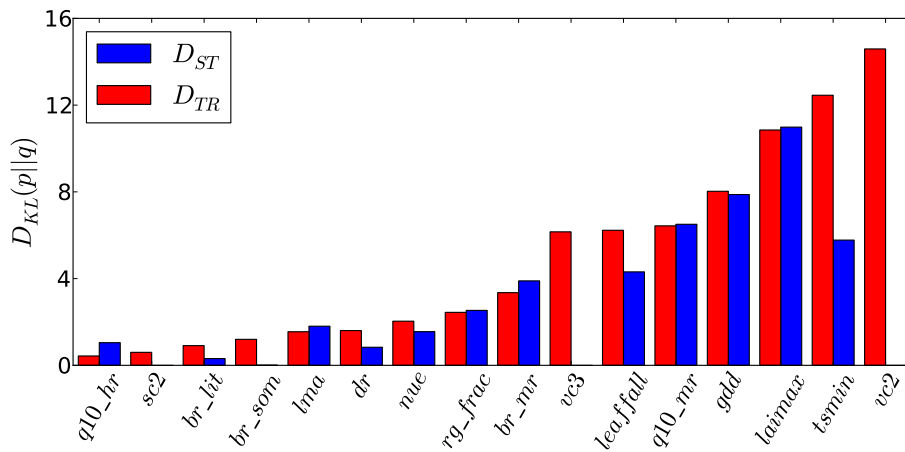

Figure 16. Kullback–Leibler divergence, $D_{KL}(p\|q)$, between prior q and posterior p densities for parameters with $D_{KL} > 0.5$ for both D_{ST} and D_{TR}.

hyperparameters, here θ:

$$p(\boldsymbol{y}|\mathcal{D}) = \int_{\theta} p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\mathcal{D})\mathrm{d}\theta. \qquad (14)$$

For the present work, $(\boldsymbol{y} - \boldsymbol{m})|\boldsymbol{\theta} \sim \epsilon_{\mathrm{m}} + \epsilon_{\mathrm{d}}$, where $y = [y_1, y_2, \ldots]$ is a N_{d}-dimensional vector with NEE predictions over a range of N_{d} days, ϵ_{m} is the model error term and ϵ_{d} is the data noise term.
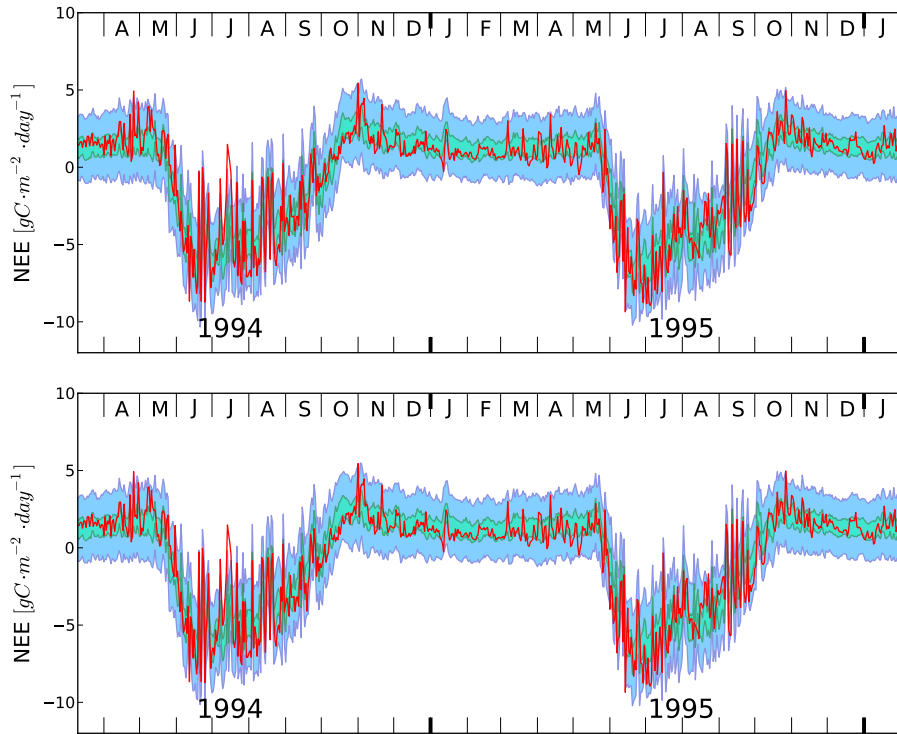
Figure 17. Marginal posterior predictive densities using the calibration results for D_{ST} (top frame) and D_{TR} (bottom frame) presented in Sect. 4. The blue regions correspond to the daily 5–95 % quantile range and the green regions to 25–75 % quantile range. The red line shows the daily NEE observations.

The 1-D marginal posterior predictive density for daily NEE values for a 2-year snapshot around 1995 are shown in Fig. 17. These densities were computed by sampling the posterior distribution of model parameters $\boldsymbol{\theta}$, i.e., by using the MCMC samples that are already available. We employ about 4000 MCMC samples, for each sample we draw 20 samples from the multivariate normal distribution $\epsilon_m + \epsilon_d$, and then add these samples to the model evaluations. These results are saved into daily bins, from which we extract several quantiles corresponding to the 1-D marginal posterior predictive density. It is worth noting that the variance of the posterior predictive distributions can also be estimated analytically as the sum of the measurement error variance and the pushed-forward variance, i.e., the variance of the output QoI with respect to posterior variability.

The top frame in Fig. 17 corresponds to D_{ST} and the bottom frame to D_{TR}. Generally, the predicted data spread covers well the observed NEE values for the entire time range. Occasional spikes can be seen outside the 5–95 % predictive band, shown in blue.

In order to quantitatively compare the predictive capability of the calibrated models for D_{ST} and D_{TR}, we adopt a probabilistic score based on the predictive cumulative distribution function (CDF). The CRPS (Gneiting and Raftery, 2007) measures the difference between the CDF of the provided data and that of the forecast/predicted data, i.e., data gener-

ated based on the posterior predictive distribution. Thus,

$$\text{CRPS}(\mathcal{F}, \mathcal{D}) = \frac{1}{N_d} \sum_{k=1}^{N_d} \int_{-\infty}^{\infty} \left(\mathcal{F}_k(y_k|\mathcal{D}) - \mathcal{H}_{\mathcal{D}_k}(y_k) \right)^2 dy_k.$$

(15)

Here, $\mathcal{F}_k(y_k|\mathcal{D})$ is the 1-D marginal posterior predictive CDF for day/component k computed using 1-D marginal posterior predictive distributions

$$\mathcal{F}_k(y_k|\mathcal{D}) = \int_{-\infty}^{y_k} p_k\left(y_k'|\mathcal{D}\right) dy_k',$$

(16)

where

$$p_k(y_k|\mathcal{D}) = \int p(\mathbf{y}|\mathcal{D}) d\mathbf{y}_{\sim k}.$$

(17)

is the 1-D marginal posterior predictive density corresponding to day k, based on $p(\mathbf{y}|\mathcal{D})$ computed via Eq. (14). Here, $d\mathbf{y}_{\sim k} = dy_1 \cdots dy_{k-1} dy_{k+1} \cdots dy_{N_d}$. The CDF of the provided data is approximated as a Heaviside function centered at the observation value $\mathcal{D}_k$ (Hersbach, 2000), $\mathcal{H}_{\mathcal{D}_k}(y_k) = 1_{y_k \geq \mathcal{D}_k}$.

We employ the posterior predictive check data presented above to compute CRPS values for both D_{ST} and D_{TR}. For

Setup	$CRPS_{prf}$	$CRPS_{prp}$	$CRPS_{psp}$	CRPSS
D_{ST}	0.16	0.91	0.66	0.33
D_{TR}	0.16	1.42	0.62	0.63

D_{ST} we obtain a value 0.67, while for D_{TR} the CRPS value is 0.60. The lower values for D_{TR} compared to D_{ST} indicate, on average, tighter marginal predictive CDF's that are better centered around the NEE data for the setup when DALEC is run for one cycle and the carbon pools are treated as parameters. This indicates a better predictive skill for D_{TR} compared to D_{ST}.

In order to measure the effect of calibration on the predictive capability of DALEC, we employ the continuous rank predictive skill score (CRPSS) (Wilks, 2011)

$$CRPSS = \frac{CRPS_{psp} - CRPS_{prp}}{CRPS_{prf} - CRPS_{prp}}, \tag{18}$$

where $CRPS_{psp}$ is the CRPS computed above based on the posterior predictive distribution, $CRPS_{prp}$ is based on the prior predictive distribution, and $CRPS_{prf}$ is the CRPS based on "perfect" predictions. For the current study, the "perfect" predictions correspond to the hypothetical case with no model error and posterior densities for model parameters centered on the NEE observations. The prior predictive distribution is defined analogous to the posterior predictive distribution in Eq. (14), with the posterior density $p(\theta|\mathcal{D})$ being replaced by $p(\theta)$, the prior density for model parameters θ.

A CRPSS value of 0 implies no improvement of the predictive skill for the calibrated model parameters compared to the predictions based on the prior information, while a value of 1 can be achieved when the posterior distribution reduces to a point and the model prediction is the same as the corresponding experimental data. The CRPS values corresponding to the prior ($CRPS_{prp}$), posterior $CRPS_{prp}$, and the ideal case $CRPS_{prf}$ are presented in Table 4 for both D_{ST} and D_{TR}. Based on the values in this table, the CRPSS for D_{TR} shows a much stronger improvement in predictive capabilities for this model setup compared to D_{ST}.

6 Conclusions

We presented uncertainty quantification (UQ) results for a process-based ecosystem carbon model. We assembled several probabilistic methodologies in a framework that tackles the connected problems of parameter estimation and forward propagation of input uncertainties. Depending on the simulation setup, the model employs either steady-state or transient assumptions, and it is driven by meteorological

data corresponding to years 1992–2006 at the Harvard Forest site. Daily net ecosystem exchange (NEE) observations were available to calibrate the model parameters and test the performance of the model.

We first discussed global sensitivity analysis (GSA) results for the complete set of input parameters. Based on their contribution to the variance, we find that different parameters have larger impacts for NEE at certain times of the year when the processes they control become important. One example is the *tsmin* parameter, which is the critical temperature at which leaf fall begins, and mainly affects NEE in October. We found that parameter interactions can also be relevant to the variability of NEE or gross primary production (GPP). Unlike NEE and GPP which are fluxes, the carbon pools, either vegetation (TVC) or soil (TSC), tend to vary more slowly and their month-to-month variability depends on a small subset of parameters. We also found that the simulation setup affects the relative importance of parameters for NEE and TSC while GPP and TVC are less sensitive to the change between steady and transient assumptions.

We then proceeded to calibrate the model parameters in a Bayesian framework using informative priors for all parameters. In this context we examined both steady and transient assumptions for the carbon model simulations. In the latter approach the initial values for the carbon pools are part of the calibration process. The discrepancy between actual and predicted NEE values was modeled as a multivariate normal distribution with constant mean and a square exponential covariance matrix. A convergence study was performed to determine the effect of covariance matrix bandwidth on the parameters of the discrepancy term. It was found that the converged correlation length does not depend on the simulation setup and that the model discrepancy for NEE data exhibits a timescale of about 1 week.

The posterior distribution of model parameters was sampled sequentially by first considering the most relevant parameters and then progressively adding less important parameters, according to GSA-based ranking. The posterior samples, obtained with a Markov chain Monte Carlo (MCMC) algorithm, exhibit significant dependencies between some of the model parameters. Comparison of posterior densities for parameters that are common to the two model setups indicate similar calibration results.

The predictive capabilities of the model, employing the parameters' posterior distribution, were assessed qualitatively through posterior predictive checks and quantitatively through continuous rank predictive score (CRPS) computations. Based on the CRPS values, the transient model setup, for which carbon pools are set as simulation parameters, performed better, in particular when compared to results based on prior predictive distributions. Given similar calibration results for the parameters common to the two configurations, we attribute the improvement in the predictive capabilities to the calibrated carbon pools in the transient model setup.

Table 5. Nomenclature.

ACM	Aggregate canopy model
CRPS	Continuous rank predictive score
CRPSS	Continuous rank predictive skill score
DALEC	Data assimilation linked ecosystem carbon
FIM	Fisher information matrix
GPP	Gross primary production
GSA	Global sensitivity analysis
MCMC	Markov chain Monte Carlo
NEE	Net ecosystem exchange
QoI	Quantity of interest
TSC	Total soil carbon
TVC	Total vegetation carbon
$D_{KL}(p||q)$	Kullback–Leibler divergence between probability densities q and p
$L_{\mathcal{D}} = p(\mathcal{D}|\boldsymbol{\theta})$	Likelihood of the data $\mathcal{D}$ for a particular instance of model parameters $\boldsymbol{\theta}$
$p(\boldsymbol{\theta}), p(\boldsymbol{\theta}|\mathcal{D})$	Prior and posterior probability densities, for model parameters $\boldsymbol{\theta}$
$p(y|\mathcal{D})$	Posterior distribution for the predicted NEE data y
$p_k(y_k|\mathcal{D})$	Marginal posterior distribution for the predicted NEE component y_k
$\mathcal{R}(X, Y)$	Distance correlation between random variables X and Y
S_i	First-order Sobol index for parameter i
S_{ij}	Joint Sobol index for parameters i and j
S_i^T	Total-order Sobol index for parameter i
$\boldsymbol{\theta}$	Vector of parameters for DALEC

The analysis presented in this paper considered a single data series at one site only. However, the Bayesian framework employed in the parameter calibrations is well-suited to deal with both heterogeneous data and multiple model setups. We are currently exploring avenues to extend this work to multi-site studies together with employing multiple data streams to better constrain the model parameters.

The Supplement related to this article is available online at doi:10.5194/gmd-8-1899-2015-supplement.

References

Barr, A., Hollinger, D., and Richardson, A. D.: CO2 Flux Measurement Uncertainty Estimates for NACP, AGU Fall Meeting, December 2009, abstract number B54A-04B, 2009.

Barr, A., Ricciuto, D. M., Schaefer, K., Richardson, A., Agarwal, D., Thornton, P. E., Davis, K., Jackson, B., Cook, R. B., Hollinger, D. Y., van Ingen, C., Amiro, B., ans M. A. Arain, A. A., Baldocchi, D., Black, T. A., Bolstad, P., Curtis, P., Desai, A., Dragoni, D., Flanagan, L., Gu, L., Katul, G., Law, B. E., Lafleur, P., Margolis, H., Matamala, R., Meyers, T., McCaughey, H., Monson, R., Munger, J. W., Oechel, W., Oren, R., Roulet, N., Torn, M., and Verma, S.: NACP Site: Tower Meteorology, Flux Observations with Uncertainty, and Ancillary Data, available at: http://daac.ornl.gov (last access: 10 June 2015) from Oak Ridge National Laboratory Distributed Active Archive Center, doi:10.3334/ORNLDAAC/1178, 2013.

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, Global Change Biol., 11, 335–355, doi:10.1111/j.1365-2486.2005.00897.x, 2005.

Campolongo, F., Saltelli, A., Sørensen, T., and Tarantola, S.: Hitchhiker's Guide to Sensitivity Analysis, in: Sensitivity Analysis, edited by: Saltelli, A., Chan, K., and Scott, E., Wiley, Chicester, 2000.

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D. M., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Wijk, M. T. V.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, Agric. For. Meteorol., 149, 1597–1615, doi:10.1016/j.agrformet.2009.05.002, 2009.

Gamerman, D.: Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Chapman & Hall, London, 1997.

Gaudinski, J. B., Trumbore, S. E., Davidson, E. A., and Zheng, S.: Soil carbon cycling in a temperate forest: radiocarbon-based estimates of residence times, sequestration rates and partitioning of fluxes, Biogeochemistry, 51, 33–69, doi:10.1023/A:1006301010014, 2000.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1996.

Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Statist. Assoc., 102, 359–378, doi:10.1198/016214506000001437, 2007.

Haario, H., Saksman, E., and Tamminen, J.: An adaptive Metropolis algorithm, Bernoulli, 7, 223–242, 2001.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems., Wea. Forecasting, 15, 559–570, 2000.

Hill, T. C., Ryan, E., and Williams, M.: The use of CO_2 flux time series for parameter and carbon stock estimation in carbon cycle research, Global Change Biol., 18, 179–193, doi:10.1111/j.1365-2486.2011.02511.x, 2012.

Kaminski, T., Knorr, W., Rayner, P. J., and Heimann, M.: Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, Global Biogeochem. Cycles, 16, 14-1–14-16, doi:10.1029/2001GB001463, 2002.

Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Consistent assimilation of MERIS FAPAR and atmospheric CO_2 into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences, 9, 3173–3184, doi:10.5194/bg-9-3173-2012, 2012.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M.: Markov Chain Monte Carlo in Practice: A Roundtable Discussion, The Amer. Statist., 52, 93–100, doi:10.1080/00031305.1998.10480547, 1998.

Knorr, W. and Kattge, J.: Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling, Global Change Biol., 11, 1333–1351, doi:10.1111/j.1365-2486.2005.00977.x, 2005.

Kucherenko, S., Tarantola, S., and Annoni, P.: Estimation of global sensitivity indices for models with dependent variables, Comput. Phys. Commun., 183, 937–946, doi:10.1016/j.cpc.2011.12.020, 2012.

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.

LeBauer, D. S., Wang, D., Richter, K. T., Davidson, C. C., and Dietze, M. C.: Facilitating feedbacks between field measurements and ecosystem models, Ecol. Monogr., 83, 133–154, doi:10.1890/12-0137.1, 2012.

Lynch, S. M. and Western, B.: Bayesian posterior predictive checks for complex models, Sociol. Meth. Res., 32, 301–335, doi:10.1177/0049124103257303, 2004.

Mahecha, M. D., Reichstein, M., Jung, M., Seneviratne, S. I., Zaehle, S., Beer, C., Braakhekke, M. C., Carvalhais, N., Lange, H., Le Maire, G., and Moors, E.: Comparing observations and process-based simulations of biosphere-atmosphere exchanges on multiple timescales, J. Geophys. Res.-Biogeosciences, 115, G02003, doi:10.1029/2009JG001016, 2010.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, Biogeosciences, 3, 571–583, doi:10.5194/bg-3-571-2006, 2006.

Raftery, A. E. and Lewis, S.: How Many Iterations in the Gibbs Sampler?, in: Bayesian Statistics 4, pp. 763–773, Oxford University Press, 1992.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), Global Biogeochem. Cycles, 19, GB2026, doi:10.1029/2004GB002254, 2005.

Ricciuto, D. M., Davis, K. J., and Keller, K.: A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty, Global Biogeochem. Cycles, 22, GB2030, doi:10.1029/2006GB002908, 2008.

Ricciuto, D. M., King, A. W., Dragoni, D., and Post, W. M.: Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length, J. Geophys. Res.-Biogeosciences, 116, G01033, doi:10.1029/2010JG001400, 2011.

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., Chen, J. M., Ciais, P., Davis, K. J., Desai, A. R., Dietze, M. C., Dragoni, D., Garrity, S. R., Gough, C. M., Grant, R., Hollinger, D. Y., Margolis, H. A., McCaughey, H., Migliavacca, M., Monson, R. K., Munger, J. W., Poulter, B., Raczka, B. M., Ricciuto, D. M., Sahoo, A. K., Schaefer, K., Tian, H., Vargas, R., Verbeeck, H., Xiao, J., and Xue, Y.: Terrestrial biosphere models need better representation of vegetation phenology: results from the North American Carbon Program Site Synthesis, Global Change Biol., 18, 566–584, doi:10.1111/j.1365-2486.2011.02562.x, 2012.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun., 145, 280–297, doi:10.1016/S0010-4655(02)00280-1, 2002.

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B., Ricciuto, D. M., and Thornton, P. E.: Dimensionality reduction for complex models via Bayesian compressive sensing, Int. J. Uncert. Quant., 4, 63–93, doi:10.1615/Int.J.UncertaintyQuantification.2013006821, 2014.

Scott, D. W.: Multivariate Density Estimation. Theory, Practice and Visualization, Wiley, New York, 1992.

Silverman, B. W.: Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

Sivia, D. S.: Data Analysis: A Bayesian Tutorial, Oxford Science, Oxford, UK, 1996.

Sobol, I. M.: Sensitivity Estimates for Nonlinear Mathematical Models, Math. Modeling and Comput. Exper., 1, 407–414, 1993.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K.: Measuring and Testing Dependence by Correlation of Distances, Ann. Statist., 35, 2769–2794, doi:10.1214/009053607000000505, 2007.

Tang, J. and Zhuang, Q.: A global sensitivity analysis and Bayesian inference framework for improving the parameter estimation and prediction of a process-based Terrestrial Ecosystem Model, J. Geophys. Res.-Atmos., 114, D15303, doi:10.1029/2009JD011724, 2009.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability, Global Biogeochem. Cycles, 21, doi:10.1029/2006GB002868, 2007.

Tjiputra, J. F., Polzin, D., and Winguth, A. M. E.: Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization, Global Biogeochem. Cycles, 21, GB1001, doi:10.1029/2006GB002745, 2007.

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, J. Geophys. Res.-Biogeosciences, 112, G02027, doi:10.1029/2006JG000367, 2007.

Urbanski, S., Barford, C., Wofsy, S. C., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J. W.: Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest, J. Geophys. Res.-Biogeosciences, 112, 1–25, doi:10.1029/2006JG000293, 2007.

White, M. A., Thornton, P. E., Running, S. W., and Nemani, R. R.: Parameterization and Sensitivity Analysis of the BIOME-BGC Terrestrial Ecosystem Model: Net Primary Production Controls, Earth Interactions, 4, 1–85, doi:10.1175/1087-3562(2000)004<0003:PASAOT>2.0.CO;2, 2000.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, Oxford, UK, 2011.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a Quercus' Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, Plant, Cell Environ., 19, 911–927, doi:10.1111/j.1365-3040.1996.tb00456.x, 1996.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biol., 11, 89–105, doi:10.1111/j.1365-2486.2004.00891.x, 2005.

Xu, T., White, L., Hui, D., and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, Global Biogeochem. Cycles, 20, GB2007, doi:10.1029/2005GB002468, 2006.

Ziehn, T., Scholze, M., and Knorr, W.: On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models, Global Biogeochem. Cycles, 26, doi:10.1029/2011GB004185, 2012.

Zobitz, J. M., Desai, A. R., Moore, D. J. P., and Chadwick, M. A.: A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC), Oecologia, 167, 599–611, doi:10.1007/s00442-011-2107-9, 2011.