



# Parameter-induced uncertainty quantification of soil N<sub>2</sub>O, NO and CO<sub>2</sub> emission from Höglwald spruce forest (Germany) using the LandscapeDNDC model

K.-H. Rahn<sup>1</sup>, C. Werner<sup>1,\*</sup>, R. Kiese<sup>1</sup>, E. Haas<sup>1</sup>, and K. Butterbach-Bahl<sup>1</sup>

<sup>1</sup>Karlsruhe Institute of Technology, Institute for Meteorology and Climate Research, Atmospheric Environmental Research, Kreuzteckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany

\* now at: Biodiversity and Climate Research Centre (BIK-F), Senckenberg Gesellschaft für Naturforschung, Senckenberganlage 25, 60325 Frankfurt am Main, Germany

Correspondence to: R. Kiese (ralf.kiese@kit.edu)

Received: 14 February 2012 – Published in Biogeosciences Discuss.: 27 April 2012

Revised: 14 August 2012 – Accepted: 10 September 2012 – Published: 17 October 2012

**Abstract.** Assessing the uncertainties of simulation results of ecological models is becoming increasingly important, specifically if these models are used to estimate greenhouse gas emissions on site to regional/national levels. Four general sources of uncertainty effect the outcome of process-based models: (i) uncertainty of information used to initialise and drive the model, (ii) uncertainty of model parameters describing specific ecosystem processes, (iii) uncertainty of the model structure, and (iv) accurateness of measurements (e.g., soil-atmosphere greenhouse gas exchange) which are used for model testing and development.

The aim of our study was to assess the simulation uncertainty of the process-based biogeochemical model LandscapeDNDC. For this we set up a Bayesian framework using a Markov Chain Monte Carlo (MCMC) method, to estimate the joint model parameter distribution. Data for model testing, parameter estimation and uncertainty assessment were taken from observations of soil fluxes of nitrous oxide (N<sub>2</sub>O), nitric oxide (NO) and carbon dioxide (CO<sub>2</sub>) as observed over a 10 yr period at the spruce site of the Höglwald Forest, Germany. By running four independent Markov Chains in parallel with identical properties (except for the parameter start values), an objective criteria for chain convergence developed by Gelman et al. (2003) could be used.

Our approach shows that by means of the joint parameter distribution, we were able not only to limit the parameter space and specify the probability of parameter values, but also to assess the complex dependencies among model

parameters used for simulating soil C and N trace gas emissions. This helped to improve the understanding of the behaviour of the complex LandscapeDNDC model while simulating soil C and N turnover processes and associated C and N soil-atmosphere exchange.

In a final step the parameter distribution of the most sensitive parameters determining soil-atmosphere C and N exchange were used to obtain the parameter-induced uncertainty of simulated N<sub>2</sub>O, NO and CO<sub>2</sub> emissions. These were compared to observational data of an calibration set (6 yr) and an independent validation set of 4 yr.

The comparison showed that most of the annual observed trace gas emissions were in the range of simulated values and were predicted with a high certainty (Root-mean-squared error (RMSE) NO: 2.4 to 18.95 g N ha<sup>-1</sup> d<sup>-1</sup>, N<sub>2</sub>O: 0.14 to 21.12 g N ha<sup>-1</sup> d<sup>-1</sup>, CO<sub>2</sub>: 5.4 to 11.9 kg C ha<sup>-1</sup> d<sup>-1</sup>). However, LandscapeDNDC simulations were sometimes still limited to accurately predict observed seasonal variations in fluxes.

## 1 Introduction

Trace gas emissions (N<sub>2</sub>O, NO and CO<sub>2</sub>) from soils of terrestrial ecosystems are highly variable in space and time due to the interplay of climatic drivers (mainly rainfall and temperature) and various ecosystem processes involved in C and N transformation and associated production and consumption

of trace gases. Therefore, quantification of the annual sink or source strength of soil greenhouse gases (GHG) is still a challenge. For sound estimates at site scale, measurements are labour and cost intensive since they should be carried out at high temporal resolution covering full annual cycles (Kiese et al., 2005; Werner et al., 2006). For that reason quantification of soil GHG emission on regional/national scale cannot solely depend on measurements, but needs to follow an integrated measuring and modelling approach. In recent years, an increasing number of biogeochemical models were tested on site scale and, after sound validation, were applied in a coupled GIS model approach for regionalisation of soil GHG emissions (Del Grosso et al., 2006; Kesik et al., 2006; Pathak et al., 2005; Li et al., 2004; Salas et al., 2007; Potter et al., 1996; Butterbach-Bahl et al., 2001; Kiese et al., 2005; Werner et al., 2007). This approach is in line with the IPCC recommendations and requirements to develop improved inventories by use of biogeochemical models. However, the so-called Tier 3 approach includes not only up-scaling of GHG emissions, but also the obligation to perform uncertainty quantification of the simulation results.

Uncertainty of model predictions can be classified into four categories: (i) uncertainty of information used to initialise and drive the model (Vrugt et al., 2008; Wikle, 2003), (ii) uncertainty of model parameters (e.g., describing specific ecosystem processes) (Vrugt et al., 2003), (iii) uncertainty of the model structure (Refsgaard et al., 2006) and (iv) accurateness of measurements (e.g., soil-atmosphere greenhouse gas exchange), which are used for model improvement and development (e.g. van Oijen et al., 2005). Uncertainty estimates in many modelling studies that investigate the soil-atmosphere exchange of trace gases only cover the assessment of uncertainty imposed by input data (e.g. Li et al., 2004; Werner et al., 2007; Winiwarter and Rypdal, 2001; Kiese et al., 2005). Due to the high complexity and large number of model parameters, work focused less on uncertainty related to model parameters as the computational demand of complex models is high and often model adaptations are required to allow application of statistical methods.

The Bayesian approach was increasingly used to quantify model parameter uncertainty on simulation results of process-based models in recent years. The Bayesian theorem was used for calibration and uncertainty assessment of parameters of dynamic process-based forest models mainly focusing on carbon turnover (van Oijen et al., 2005; Svensson et al., 2008; Klemetsson et al., 2008) and more recently also for parameters involved in production, consumption, transport and emissions of soil GHGs (e.g. Lehuger et al., 2009). To our knowledge van Oijen et al. (2011) is the only study so far comparing four process-based biogeochemical forest models within a Bayesian model comparison framework. In contrast to such a model inter-comparison, the aim of this study is to provide deeper insights into the individual parameter uncertainty and calibration of the model LandscapeDNDC and the subsequent uncertainty of simulated trace gas

exchange. The parameter distribution, which was estimated after an objective multi-chain convergence check, was additionally tested on a validation dataset.

LandscapeDNDC is a process-oriented biogeochemical model, which simulates the biosphere-atmosphere exchange of greenhouse gases on the basis of the simulation of all major ecosystem C and N cycling processes (Haas et al., 2012; Werner et al., 2012).

We used a time series covering 10 yr of soil-atmosphere trace gas fluxes as observed continuously in sub-daily time resolution at the Höglwald spruce forest, Germany (e.g. Butterbach-Bahl et al., 2002; Wu et al., 2010) to assess the model parameter uncertainty of the LandscapeDNDC model.

Results of the Bayesian calibration approach can be used to gain insights into the complex parameter dependencies, to identify weaknesses in process descriptions and to narrow the range of likely model parameter values, which finally reduces uncertainty of the simulation results.

## 2 Model description and model parameter selection

The LandscapeDNDC model applied in this study is a derivative of the DNDC model family (Li et al., 1992, 2000) and was further developed from the MoBiLE model framework (Grote et al., 2009, 2011). LandscapeDNDC incorporates functions of DNDC (agricultural sites) and PnET-N-DNDC/Forest-DNDC (forest sites), which were initially set up to predict soil carbon and nitrogen biogeochemistry with a specific focus on the simulation of soil N trace gas emissions (Li et al., 2000; Stange et al., 2000; Butterbach-Bahl et al., 2001; Kiese et al., 2005; Kesik et al., 2005; Werner et al., 2007). LandscapeDNDC integrates different modules for describing soil environmental conditions (temperature, moisture, pH, nutrient availability and anaerobic volume fractions), soil-chemistry integrating microbial C and N turnover processes (mineralisation, nitrification and denitrification) and associated C and N trace gas emissions (e.g., N<sub>2</sub>O, NO and CO<sub>2</sub>) as well as vegetation dynamics (Grote, 2007). It also offers a flexible initialisation of vegetation and soil properties and efficient multi-site calculations that ease regional applications as well as sensitivity and uncertainty studies (Haas et al., 2012).

Each module includes parameters derived from physical and chemical principals and laboratory measurements. In this study, we focus on the analysis of parameter-induced uncertainty quantification stemming from the soil-chemistry module describing all soil processes relevant for C and N trace gas production, consumption and transport, being crucial for the simulation of soil-atmosphere GHG exchange. Here, we do not consider model parameters of other modules e.g., for plant growth and soil water cycling modules in order to reduce complexity and degrees of freedom and to increase the efficiency of the calibration process. However, these modules were tested and calibrated in recent studies (e.g. Kiese et al., 2011) and are run using default parameters.

The soil-chemistry sub-module in total holds 67 parameters, mostly describing biological kinetics of nutrient turnover and transformation by growth and death of different types of microbes (e.g., nitrifiers and denitrifiers). Parameter values are generally derived from laboratory measurements and expert knowledge, if detailed information is not available. This introduces different levels of uncertainty, which need to be quantified and requires calibration.

The model parameters can only be estimated and optimised by an inverse calibration technique (cf. Vrugt et al., 2003), which compares model simulation output by using randomly selected model parameter vectors with measured observations. The observational data used was collected at the Höglwald spruce forest, Germany, covering the years 1994 to 1997 (Papen and Butterbach-Bahl, 1999; Gasche and Papen, 1999) and 2002 to 2003. The remaining observation period (years 2004 to 2007) was used for validation purpose and finally for assessing the prediction uncertainty.

Each parameter included into the uncertainty analysis adds a new dimension in the parameter space. Therefore, computational cost rises tremendously with the increasing number of parameters while efficiency of the calibration technique decreases. Furthermore, correlations among parameters become more likely by increasing the number of parameters. This subsequently leads to slower convergence rates (requiring additional iterations), as parameter vectors, which do not comply with these relations, are less likely to be accepted by the Bayesian algorithm (cf. Gilks et al., 1996). Additionally, a higher degree of freedom exist, i.e., parameter configurations producing similar outputs may not be unique. To avoid these obstacles we used a sensitivity analysis (Saltelli, 2008) developed by Morris (1991) prior to the Bayesian calibration method. This helps to restrict the analysis to the most influential parameters and to avoid over-fitting effects.

The method introduced by Morris is an efficient tool for parameter screening, since it can easily be implemented and computational demands are low (van Oijen et al., 2011). The method varies parameter values and finally produces a ranking of the model parameters based on their impact on the simulated model output of C and N trace gas emissions and soil moisture.

This procedure divides each parameter range in  $n$  (here  $n = 6$ ) equidistant levels, starts with a random parameter vector using these levels and randomly changes one parameter after another to one of the other levels (1 iteration). Differences in model output are stored and used to rank the model parameters according to their influence on the simulation output. Since the trajectory of parameter changes per iteration is randomly selected  $m$  times (here  $m = 50$ ), the method spans the parameter space better than a “one-parameter-at-a-time approach” (see Hamby, 1994). The model parameters, which produce largest differences (i.e., having highest sensitivity on the output variable), are regarded as the most influential ones.

We initialised and ran the model with specific site information (soil, vegetation and climate) of the Höglwald spruce forest to identify the most sensitive parameters of LandscapeDNDC affecting soil C and N fluxes. This approach does not require a comparison of simulated emission to measurements, since the sensitivity analysis only focuses on parameter-induced changes of model output. Parameter sensitivities were calculated separately for the output variables of soil  $\text{N}_2\text{O}$ , NO and  $\text{CO}_2$  emissions, which finally resulted in three different parameter-ranking lists. We selected the first 20 most influential parameters of any list, thereby considering the trade-off between over-parameterisation and under-representing significant processes. Due to close linkage of C and N cycling and in particular NO and  $\text{N}_2\text{O}$  emission there was a good overlap of the most sensitive parameters. This led to an overall selection of 26 parameters (see Table 1).

We regressed the stored model output (a) to all parameters and (b) to the reduced parameter subset and compared the adjusted coefficient of determination  $\bar{R}^2$  of both linear regressions (cf. van Oijen et al., 2011) to evaluate, whether the reduced parameter set accounts for most of the models behaviour. The results show that for  $\text{N}_2\text{O}$  and  $\text{CO}_2$  more than 90 % and for NO 65 % of the models linear behaviour is explained by the subset of the parameters. We regard these numbers to be sufficient for continuing the Bayesian calibration approach with the restricted parameter set and at the same time assure a balance with calibration efficiency, which will be reduced when introducing more parameters as already stated before. Following the selection of the most sensitive model parameters, the joint parameter distribution given the data was estimated by means of a Bayesian calibration. From this distribution, parameter values can be sampled to perform simulation runs and finally address the frequency distribution of simulation results. See Fig. 1 for an illustration of the workflow.

### 3 Bayesian calibration

In a standard frequency approach the parameter value is not regarded as a random variable. The used parameter value is either the true value or it is not (see Ellison, 1996). Therefore, a Bayesian approach is needed (Clark, 2005; van Oijen et al., 2005; Klemedtsson et al., 2008; Gelman et al., 2003; Reinds et al., 2008; Lehuger et al., 2009) since it models the parameter vector  $\theta$  as a random vector, which allows a direct quantification of the probability of a certain parameter realisation/range.

The probability density of a parameter value  $\theta$  given the measurement  $D$  (posterior) is:

$$p(\theta|D). \quad (1)$$

**Table 1.** Selected parameters being most influential for simulating soil-atmosphere trace gas fluxes (N<sub>2</sub>O, NO and CO<sub>2</sub>) with LandscapeDNDC.

Parameter	Description
D_N2O	effective N <sub>2</sub> O diffusion constant [m <sup>2</sup> h <sup>-1</sup> ]
D_NO	effective NO diffusion constant [m <sup>2</sup> h <sup>-1</sup> ]
DRF	scaling factor for decomposition rate constants of SOM
EFFAC	partitioning of CO <sub>2</sub> and DOC production during microbial decomposition of organic matter
FNO3_U	fraction of microbial N-uptake as (NO <sub>3</sub> )
FRC	factor regulating microbial death depending on the availability of very labile and labile carbon
FTRANS	factor regulating microbial nitrate immobilisation and direct re-mineralisation to NH <sub>4</sub>
KCRB_L	decomposition constant of labile dead microbial biomass
KHDC_L	decomposition constant of labile humads
KHDC_R	decomposition constant of recalcitrant humads
KM.O2	factor regulating splitting of DOC and CO <sub>2</sub> during decomposition of SOM depending on O <sub>2</sub> concentration
KMDC.DOC	factor for half optimum content of doc in soil solution for denitrifier activity [kg Cha <sup>-1</sup> ]
KMDC.N	factor for half optimum content of nitrogen in soil solution for denitrifier activity [kg N ha <sup>-1</sup> ]
KMM.DOC	factor regulating growth of microbes in dependency of DOC substrate
KMNO2	factor regulating NO <sub>2</sub> to NO <sub>3</sub> conversion depending on NO <sub>2</sub> concentration during nitrification
KN2O	loss rate of N <sub>2</sub> O during nitrification
KNO	loss rate of NO during nitrification
KRCL	decomposition constant for labile litter pool
KRCR	decomposition constant for recalcitrant litter pool
MICRRESP	factor regulating CO <sub>2</sub> production during microbial metabolism in dependency of microbial C/N ratio
NH4_DENIMAX	maximum fraction of NH <sub>4</sub> available for auto- and heterotrophic nitrification
PERTL	fraction of labile litter, which can be reallocated into deeper soil layers
PERTR	fraction of recalcitrant litter, which can be reallocated into deeper soil layers
PERTVL	fraction of very labile litter, which can be reallocated into deeper soil layers
PSL.SC	depth dependent factor for reallocation of organic matter into deeper soil layers
SRB	fraction of labile dead microbial biomass

By using Bayes theorem, the posterior is proportional to the product of the likelihood  $p(D|\theta)$  and the prior density  $p(\theta)$ :

$$p(\theta|D) \propto p(D|\theta) \cdot p(\theta). \quad (2)$$

The prior, describing the a priori knowledge on parameters, is determined by using literature data and biogeochemical principles to address the most likely parameter value and to constrain the range of a parameter. We use an uninformed prior (uniform distribution) ranging between provided minima and maxima for the given parameter as derived from expert knowledge or laboratory and field experiments. The likelihood, the only unknown term, describes the probability of a data realisation for a particular parameter vector.

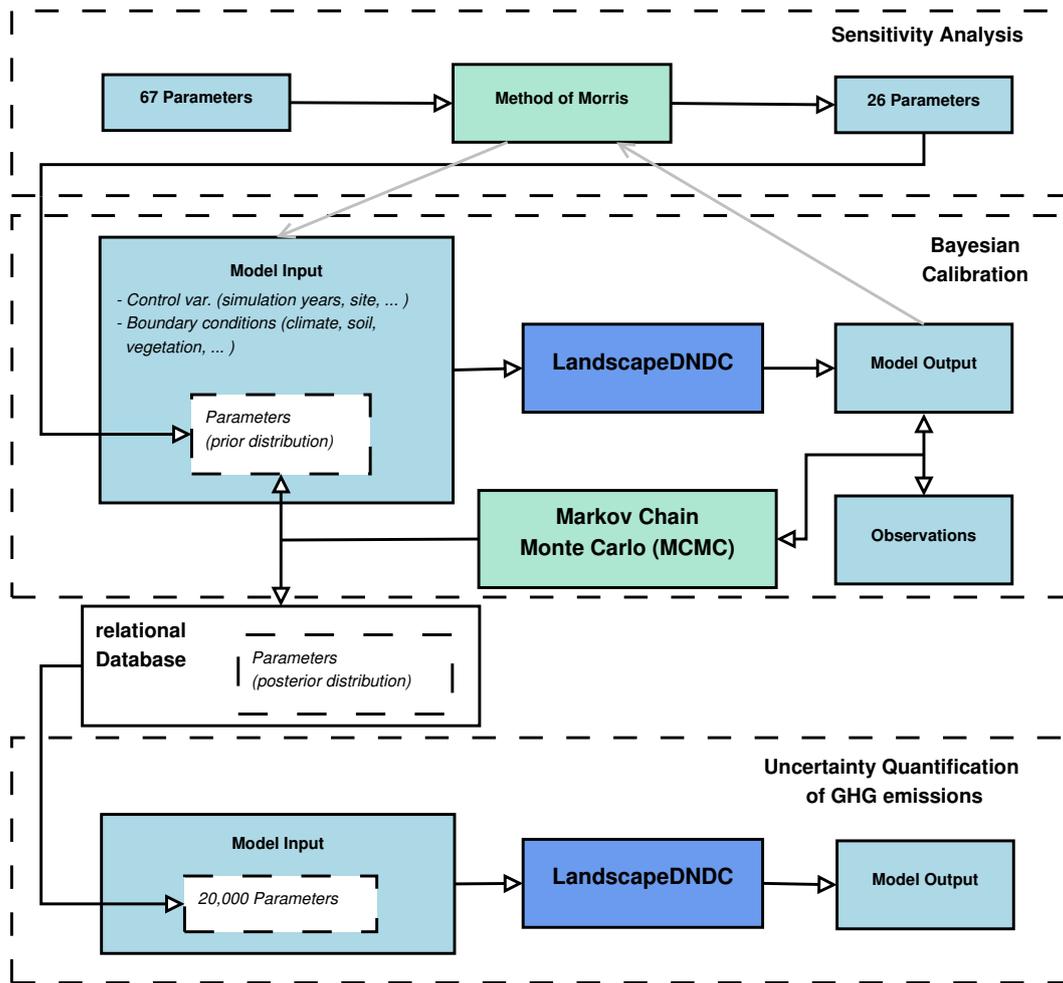
We assume the difference  $D - M$  between data  $D$  and model  $M$  to be normal distributed, hence the likelihood is (van Oijen et al., 2005):

$$p(D|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(D-M(\theta))^2}{2\sigma^2}} \quad (3)$$

Since this term cannot be solved analytically, a Metropolis algorithm (Metropolis et al., 1953) generates a Markov chain,

which samples from the posterior distribution after convergence of the chain (see next section for convergence criteria).

Although the applied LandscapeDNDC model was run on daily time-step, for the Bayesian calibration, daily simulated as well as measured trace gas fluxes were aggregated to weekly means in order to avoid that minor temporal lags (1–2 days) between daily measured and simulated peak emissions penalise likelihood calculations. Using this approach (see also van Oijen et al., 2011) we find a balance for increasing acceptance-rates resulting in a more conservative estimation of the parameter uncertainties while still representing the temporal dynamic of C and N trace gas emissions. Furthermore, reduction of data by using weekly rather than daily fluxes, helps to prevent asymptotic collapse of the posterior and an overreliance on the information of the data (Arhonditis et al., 2008). Although within the calibration procedure, model performance evaluation was based on weekly aggregated data the uncertainty quantification was done on a daily simulation time step. We also tested Bayesian calibration of parameters by using monthly aggregated data. However, this resulted in a significant flattening of the daily simulated



**Fig. 1.** Schematic view of the workflow for assessing the uncertainty of simulated soil GHG emissions while using LandscapeDNDC. After reduction to influential parameters by means of a sensitivity analysis, the distribution of the model parameters was estimated using a Bayesian calibration. Subsequently, an uncertainty quantification of simulated emissions was carried out using 20 000 samples out of the 533 000 post burn-in realisations of the parameter distribution stored in a relational database.

emission pattern which was not the case with the parameter set of weekly aggregated data (Fig. 6).

In order to increase computation efficiency, we run the model in parallel for the six simulated calibration years on a High Performance Computing (HPC) Linux cluster.

### 3.1 Criteria to define convergence while using a multi-chain approach

As it is not possible to draw any statistical inference from the sampled parameter vectors if the Markov chain has not converged (Gilks et al., 1996), we used four independent Markov chains (differing only in the individual parameter starting points) and tested for convergence at each iteration step. When convergence was reached (end of “burn in phase”), the previous parameter samples were discarded and all following data were included in the further analysis.

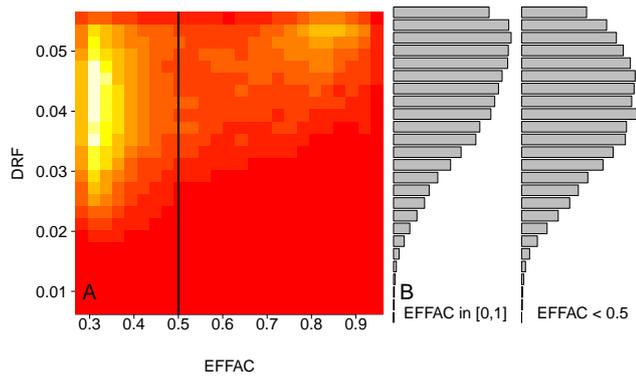
To quantify convergence, Gelman et al. (2003) introduced the measure  $\hat{R}$  which compares the variances of each chain (within sequence variance, Eq. 4) to the joint variance of all chains (between sequence variance, Eq. 5)

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2, \text{ where } s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{i,j} - \bar{\psi}_{\cdot,j})^2 \quad (4)$$

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\psi}_{\cdot,j} - \bar{\psi}_{\cdot\cdot})^2, \quad (5)$$

where  $\bar{\psi}_{\cdot,j} = \frac{1}{n} \sum_{i=1}^n \psi_{i,j}$ ,  $\bar{\psi}_{\cdot\cdot} = \frac{1}{m} \sum_{j=1}^m \bar{\psi}_{\cdot,j}$ .

In the process of convergence the measure  $\hat{R} = \sqrt{\frac{n-1}{n} + \frac{B}{nW}}$  approaches from values  $>1.0$  to 1.0. As  $\hat{R}$  is not expected



**Fig. 2.** (A) Heat-map of 2-dimensional marginal distribution of EFFAC and DRF (decomposition rate factor), the brighter the polygons, the higher the posterior value. (B) histogram of DRF using all values and histogram of DRF using only values of DRF, where  $\text{EFFAC} < 0.5$ .

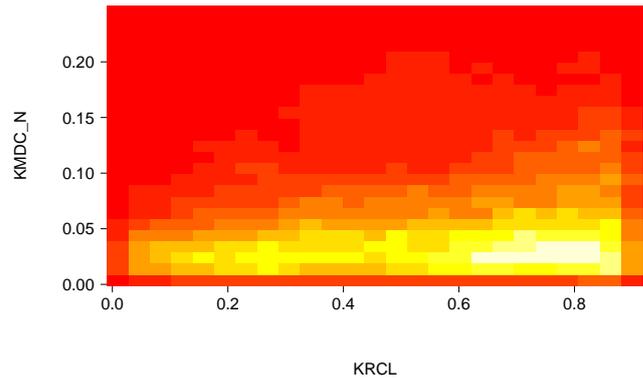
to reach exactly 1.0, a threshold of 1.2 is introduced as the acceptance threshold (Kass et al., 1998).

By using four chains, our implementation spreads the model to 24 CPUs (4 chains  $\times$  6 separate simulation years) using the Message Passing Interface (MPI). After 1000 iteration steps the Gelman/Rubin statistic was calculated and continuously updated until convergence (according to  $\hat{R}$ ) of chains. In our setup, burn-in of all parameters was completed after 31 656 iteration steps. After a visual inspection of the marginal distributions, we decided to continue the Markov chain, as one parameter (EFFAC) showed a bimodal distribution for two chains, but only one mode for the remaining chains. After additional 133 000 iterations the marginal posterior distribution of EFFAC had the same shape for all four chains. The Gelman Rubin Statistic was at the same time well below 1.1 for all parameters.

The acceptance-rates of the four chains ranged from 14.1 % to 15.8 % (using a step-width of 0.04). These are reasonable values taking into account the large dataset (6 yr of data in daily time resolution and 3 target variables:  $\text{CO}_2$ ,  $\text{N}_2\text{O}$  and  $\text{NO}$ ) and, therefore, a rather strict rejection step due to a narrow-shaped posterior (Arhonditsis et al., 2008; Clark, 2005; Rahn et al., 2011).

### 3.2 Effective data storage

The study design and computational setup lead to substantial amounts of data, which need to be efficiently handled within subsequent data analysis. For that reason a interface to a relational database was developed using Structured Query Language (SQL) which warranted a concurrent access and high data integrity.



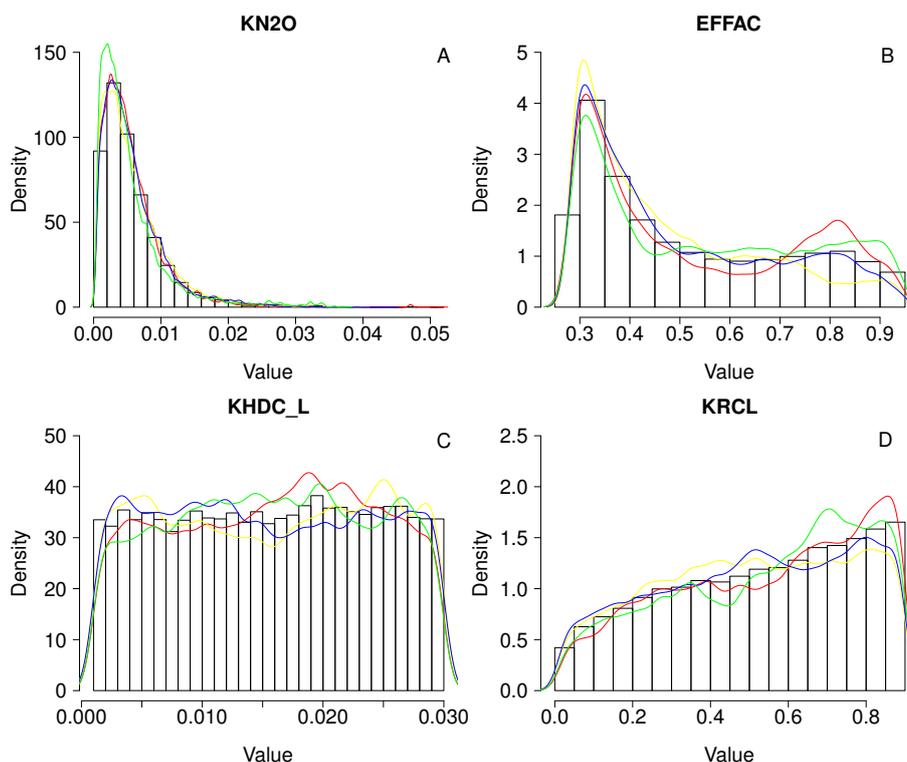
**Fig. 3.** Heat-map of 2-dimensional marginal distribution of KRCL (decomposition constant for labile litter pool) and  $\text{KMDC}_N$  (factor for half optimum content of nitrogen in soil solution for denitrifier activity). Higher values of KRCL lead to a wider range of  $\text{KMDC}_N$ .

## 4 Estimating simulated soil GHG flux distributions

In a second step the posterior distribution of the 26 parameters was used to quantify the uncertainty of LandscapeDNDC simulations for soil  $\text{N}_2\text{O}$ ,  $\text{NO}$  and  $\text{CO}_2$  emissions of the Höglwald Forest spruce site (see Fig. 1). For this, we used a total of 20 000 posterior-parameter vectors (posterior-samples) by selecting every 26th parameter vector out of the 532 000 posterior-parameter samples of the four chains (133 000 for each chain) until 20 000 parameters were taken. We, thereby, reduced dependencies between parameter vectors of consecutive iterations (Kass et al., 1998; Toft et al., 2007), which arose as each parameter vector of the posterior distribution had been taken dependent on its predecessor during the calibration process.

Following the selection of the posterior-samples, we executed LandscapeDNDC with the parameter realisations for the calibration set (years 1994 to 1997 and 2002 to 2003) and an independent validation dataset (years 2004 to 2007). As a result, we obtained distributions (including associated uncertainty) of simulated soil  $\text{N}_2\text{O}$ ,  $\text{NO}$  and  $\text{CO}_2$  emissions.

The root-mean-squared error (RMSE) is used to quantify the difference between measurements and simulations. Therefore, we defined the distance of measurements to the distribution of the simulations as the minimum of the distances between the measurements and the two boundaries of the credible interval or 0 whenever the measurement is within the range of the credible interval. The RMSE of the best simulation ( $\text{RMSE}(\theta_{\text{MAP}})$ ) is calculated using the common definition.



**Fig. 4.** Four typical histograms of marginal parameter distributions. The coloured density lines of two right-skewed (KN2O: loss rate of N<sub>2</sub>O during nitrification and EFFAC: describing the partitioning of CO<sub>2</sub> and DOC production during microbial decomposition of organic matter), a flat (KHDC\_L: decomposition constant of labile humads pool) and a left-skewed distribution (KRCL: decomposition constant of labile litter pool) were done by post burn-in samples of each individual chain, whereas the histograms are plotted using post burn-in samples of all chains.

## 5 Results

### 5.1 Posterior parameter distribution

An estimate for the posterior distribution of the 26 most sensitive LandscapeDNDC parameters for simulation of soil N<sub>2</sub>O, NO and CO<sub>2</sub> emissions was obtained by using Bayesian calibration technique and initial information on the likely range of the selected parameters. To illustrate common features of the obtained marginal posterior distributions, we present a subset of four model parameters (see Fig. 4 and Table 2). For each marginal histogram, 4 × 133 000 post burn-in chain steps were used.

In the first histogram (Fig. 4a) the marginal distribution of the parameter KN2O, the loss rate of N<sub>2</sub>O during nitrification, is displayed. The prior parameter uncertainty ( $SD_{\text{prior}} = 0.026$ ) was reduced substantially ( $SD_{\text{post}} = 0.005$ ) and the most probable value of the right-skewed distribution is in a narrow region between 0.002 and 0.008.

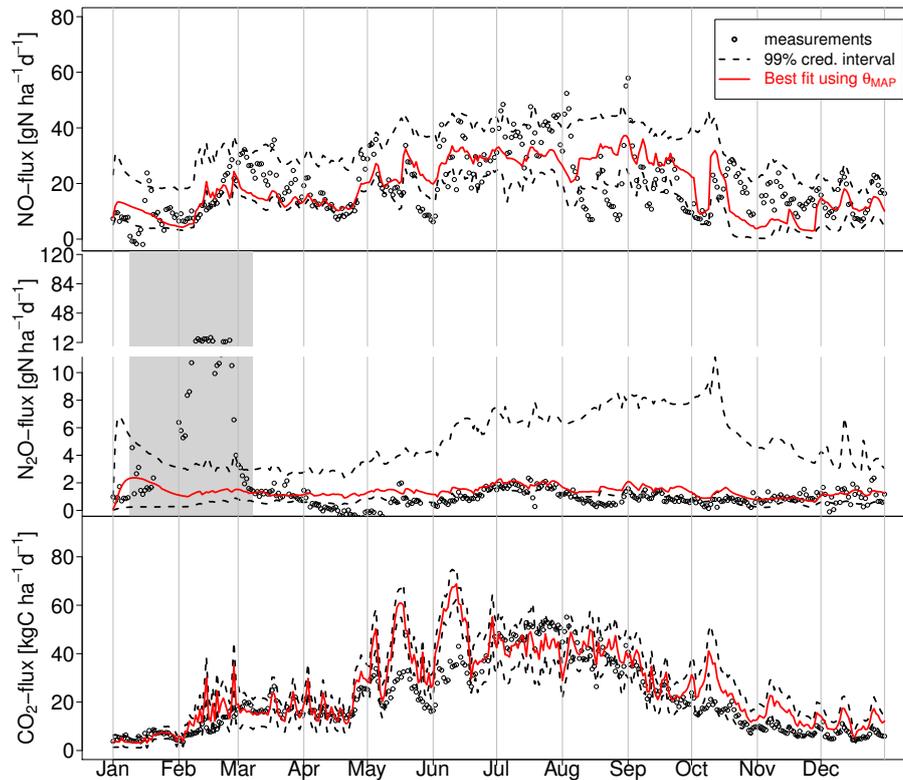
After the first 30 000 iterations, the marginal distribution of the parameter EFFAC (describing the partitioning of CO<sub>2</sub> and DOC production during microbial decomposition of organic matter) showed a bimodal shape for two out of the four chains. However, after 165 000 iterations the chains fi-

nally all sampled from the right-skewed distribution shown in Fig. 4b. As a result, the convergence rate (time needed for convergence to the posterior distribution), was low and consequently the number of iterations was high.

Figure 4c displays the posterior distribution of the parameter KHDC\_L. This parameter is the decomposition constant for the labile humads pool (death microbial biomass). For this model parameter the posterior distribution is flat, i.e., all values across the explored range are of similar probability. Here, the uncertainty of the initial parameter could not be reduced significantly by the Bayesian calibration and only values approaching zero are less likely than others.

An example for a left-skewed distribution of a parameter is given in Fig. 4d, in this case of KRCL. KRCL is the decomposition constant for the labile litter pool. Although there is a tendency for higher values, smaller values can still occur depending on the values of the other 25 parameters. In conclusion, the uncertainty of the parameter KRCL is reduced, however, not as much as compared to KN2O.

A correlation analysis for the 26 selected parameters revealed for most pair-wise constellations no relevant correlations. This might be due to the large number of sampling points over the entire parameter space (see Fig. 9). Higher correlations in absolute appeared only between KMNO2



**Fig. 5.** Simulated fluxes (calibration set) versus measurements of NO, N<sub>2</sub>O and CO<sub>2</sub> fluxes at the spruce site of the Höglwald forest in the year 1997. The grey box highlights pulse emissions of N<sub>2</sub>O during soil freeze-thaw events.

(Michaelis-Menten constant for NO<sub>2</sub> to NO<sub>3</sub> conversion during heterotrophic nitrification) and D\_NO (effective NO diffusion constant) with a correlation of  $-0.71$ , and between EFFAC and FNO3\_U (fraction of microbial N-uptake as NO<sub>3</sub>) with a correlation of  $0.51$ . All other correlations were in the range of  $\pm 0.40$ , most of them between  $\pm 0.25$  (Fig. 8).

However, that does not fully exclude any relationship between parameters, since they are often of nonlinear character. Figure 2 shows that limiting the values of EFFAC to values  $< 0.5$  leads to a more bell shaped distribution of the parameter DRF (scaling factor for decomposition rate constants of SOM) around the value  $0.042$  (correlation between EFFAC/DRF =  $0.32$ ). At the same time smaller values of FTRANS (factor regulating microbial nitrate immobilisation and direct re-mineralisation to NH<sub>4</sub>), FNO3\_U, KRCR (decomposition constant for recalcitrant litter pool) and KMM\_DOC (Michaelis-Menten constant regulating growth of microbes in dependency of DOC substrate) become more likely, whereas for other parameters like MICRRESP (factor regulating CO<sub>2</sub> production during microbial metabolism in dependency of microbial C/N ratio), FRC (factor regulating microbial death depending on the availability of very labile and labile carbon) and KN2O (loss rate of N<sub>2</sub>O during nitrification) higher values occur more often, thus, get more likely. That also shows that restricting some parameters to a range

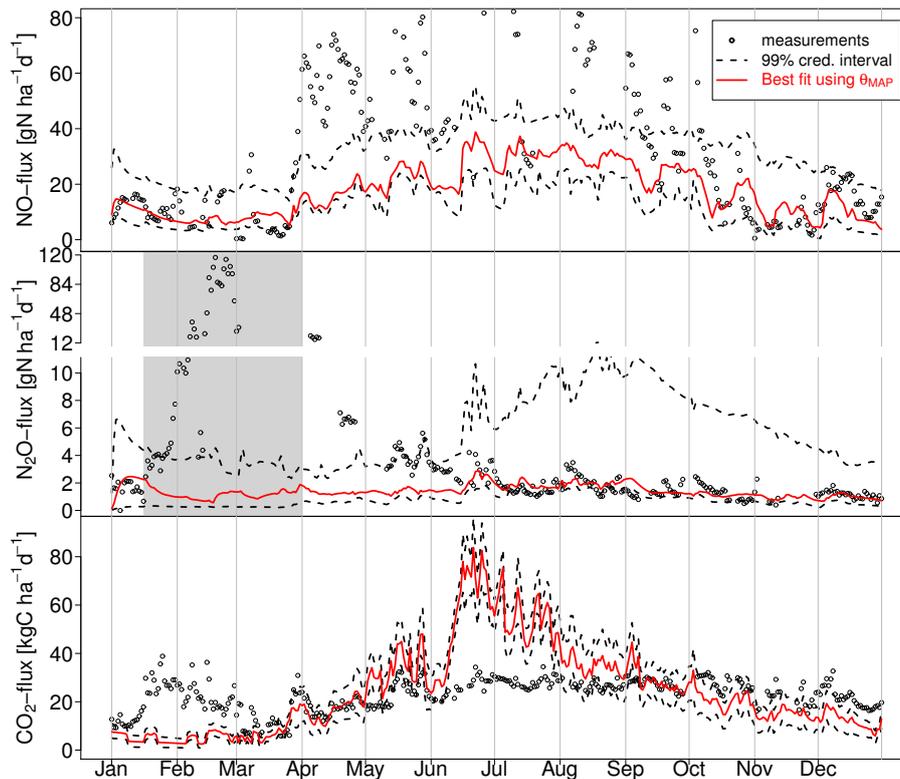
of their most likely values can narrow the range of likely values of other parameters.

The heat map presented in Fig. 3, shows the relationship between KRCL and KMDC\_N. While the correlation between the two parameters is low ( $= 0.06$ ), one can see that lower values of KRCL restrict the range of KMDC\_N to lower values. To capture all dependencies (compare Fig. 9) when estimating the distribution of model simulations, it is straightforward to use samples of the joint posterior parameter distribution, as the whole structure of parameter dependence is fully included.

## 5.2 Uncertainty quantification of soil-atmosphere gas emissions at Höglwald forest (1994 to 1997, 2002 to 2003 and 2004 to 2007)

### 5.2.1 Calibration set

In general, most of measured trace gas emissions of N<sub>2</sub>O, NO and CO<sub>2</sub> are within or close to the range of the simulated 99 % credible interval (cf. Gilks et al., 1996) (see for example Fig. 5. RMSE values for each year and each soil-atmosphere flux are presented in Table 4). Based on the evaluation criteria, LandscapeDNDC was able to correctly simulate cumulative N<sub>2</sub>O and NO emissions in five and six out of six years, respectively (see Table 3). In two out of three



**Fig. 6.** Simulated fluxes (validation set) versus measurements of NO, N<sub>2</sub>O and CO<sub>2</sub> fluxes at the spruce site of the Höglwald forest in the year 2006. The grey box highlights pulse emissions of N<sub>2</sub>O during soil freeze-thaw events.

years, cumulative CO<sub>2</sub> observations were located within the simulated CO<sub>2</sub> ranges. Comparatively high NO emissions (> 60 gNha<sup>-1</sup> d<sup>-1</sup>) measured in the summers of 1996 and 2003, however, could not be reproduced by LandscapeDNDC (model simulations underestimated fluxes in summer periods by at least 22 % and 33 %, respectively).

Seasonal dynamics of NO measurements were reproduced for the years 1994, 1997 and partly for 2002, which resulted in low RMSE values for the credible interval (RMSE(CI): 2.40 to 3.20 gNha<sup>-1</sup> d<sup>-1</sup>) and when using the maximum posterior parameter vector  $\theta_{\text{MAP}}$  (RMSE( $\theta_{\text{MAP}}$ ): 6.66 to 9.20 gNha<sup>-1</sup> d<sup>-1</sup>). Although in most of the remaining years the magnitude of measurements and simulations is similar, the temporal dynamic could not always be clearly reproduced.

N<sub>2</sub>O simulations especially suffer from the inability of the actual LandscapeDNDC version to simulate freeze-thaw pulse emissions (Papen and Butterbach-Bahl, 1999; Butterbach-Bahl et al., 2002; Wolf et al., 2010) in 1995, 1996, 1997 and 2003 (RMSE(CI): up to 16.10 gNha<sup>-1</sup> d<sup>-1</sup>). Therefore, following simulation to measurement comparisons of N<sub>2</sub>O were restricted to periods being unaffected by freeze-thaw events. Nevertheless, cumulative statistics and RMSE statistics can be compared with or without freeze-thaw events in Tables 3 and 4. One can see that the RMSE is strongly reduced when neglecting frost-thaw emissions

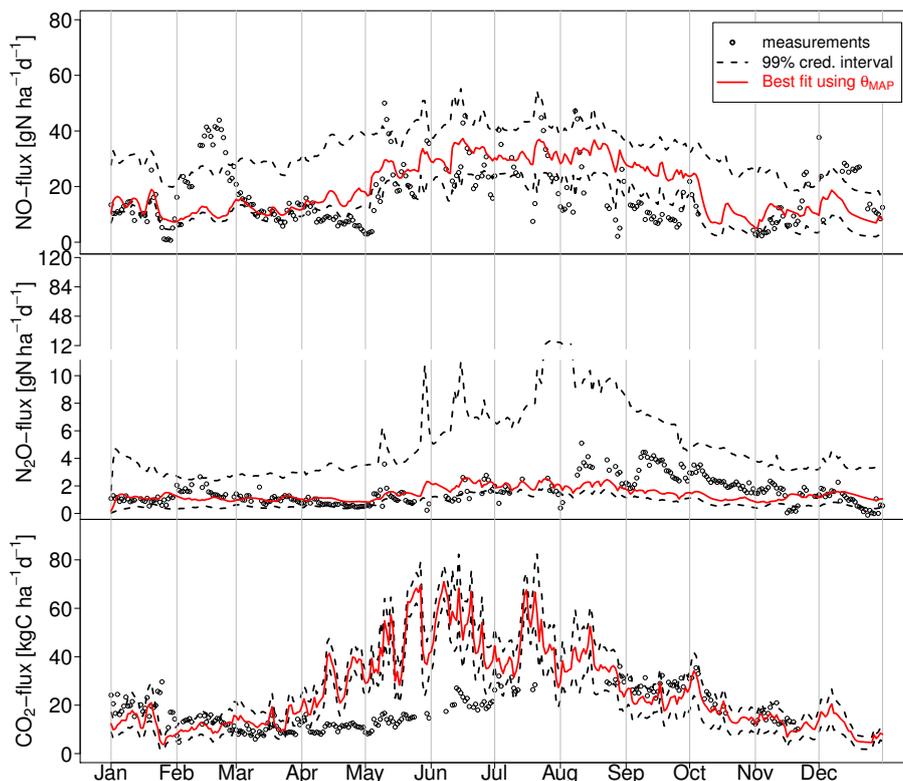
(e.g., RMSE(CI) reduced from 16.10 to 7.92 in 1996 and from 2.56 to 0.42 in 1997). Peak emissions of N<sub>2</sub>O (> 10 gNha<sup>-1</sup> d<sup>-1</sup>) in August 2002 could also not be reproduced by the model, although the model could comprehend the general increase of N<sub>2</sub>O emissions in the beginning of August (up to 7 gNha<sup>-1</sup> d<sup>-1</sup>).

CO<sub>2</sub> emissions were underestimated by at least 19 % and 7 % during August to November in 1995 and 1996. From May to June 1997, they were overestimated by at least 25 %. Note that only 1004 CO<sub>2</sub> observations were used for calibration, compared to 1890 and 2075 values for NO and N<sub>2</sub>O. Thus, CO<sub>2</sub> emissions were underweighted by a factor of approx. 0.5 in the calibration process.

## 5.2.2 Validation set

To independently validate the behaviour of the parameterisation, we simulated soil-atmosphere trace gas emissions in Höglwald for 2004 until 2007, i.e., for a time period, which has not been used for the calibration of LandscapeDNDC. The parameterisation of the model includes the same posterior-samples that have been used to simulate the emissions of the calibration set (1994 to 1997, 2002 to 2003) and to visualise model uncertainty.

For the validation set, LandscapeDNDC produced comparable results as for the calibration set. Cumulative NO



**Fig. 7.** Simulated fluxes (validation set) versus measurements of NO, N<sub>2</sub>O and CO<sub>2</sub> fluxes at the spruce site of the Höglwald forest in the year 2007.

emissions were covered by the 99 % credible interval in three out of four simulated years and in three years with regard to simulated N<sub>2</sub>O emissions. Soil CO<sub>2</sub> emissions could be reproduced for all four years. High N<sub>2</sub>O fluxes in 2006 were underestimated by at least 8 %. When excluding the freeze-thaw event in 2006 the simulated and measured cumulated N<sub>2</sub>O emissions were in the same range. The averaged RMSE values of the validation set (0.91 gNha<sup>-1</sup> d<sup>-1</sup>) were even lower than the averaged RMSE values of the calibration set (1.70 gNha<sup>-1</sup> d<sup>-1</sup>).

The large discrepancy between soil NO simulations and measurements in 2006 (RMSE = 18.95 gNha<sup>-1</sup> d<sup>-1</sup>) is due to underestimation of the simulated fluxes mainly in the vegetation period (April to October by at least 33 %).

Soil CO<sub>2</sub> fluxes were mainly overestimated by the model in summer (e.g., 45 % May to July 2005; 54 % June to August 2006; 73 % April to August 2007). Model deviations from measured soil CO<sub>2</sub> fluxes are also obvious during wintertime in 2005 and 2006 (see Fig. 6).

## 6 Discussion

Our work shows that the Bayesian calibration approach can successfully be implemented to estimate the posterior parameter distribution of a complex sub-module of a biogeochemi-

cal model used for simulating soil N<sub>2</sub>O, NO and CO<sub>2</sub> fluxes at a spruce site of the Höglwald forest, Germany. The applicability of the illustrated method to complex ecological models was also demonstrated in previous studies (e.g. van Oijen et al., 2005; Svensson et al., 2008; Klemedtsson et al., 2008; Lehuger et al., 2009).

Bayesian calibration reduced the prior uncertainty (by up to 82 %) for 15 out of 26 parameters for simulating soil-atmosphere exchange of C and N trace gases. For the remaining 11 parameters the calibration process achieved no significant reduction in parameter uncertainty. The flat shape of the distribution of these 11 parameters occurred because different parameter constellations can lead to similar model output. The underlying reasons for that cannot be further specified, as the parameter space is 26-dimensional and small changes in high-sensitive parameters may be compensated by changes of (many or all) remaining parameters.

Correlations between parameters were hardly found since only for two cases parameter correlation was significant (51 % and -71 %). This is most likely due to the fact, that interactions between parameters in a 26-dimensional parameter space are more complicated than linear relations between only 2 parameters, which was shown exemplarily for the parameters EFFAC and DRF. For some of the parameters (especially for those with a flat marginal distribution) this finding might also point to an overparametrisation of the model.

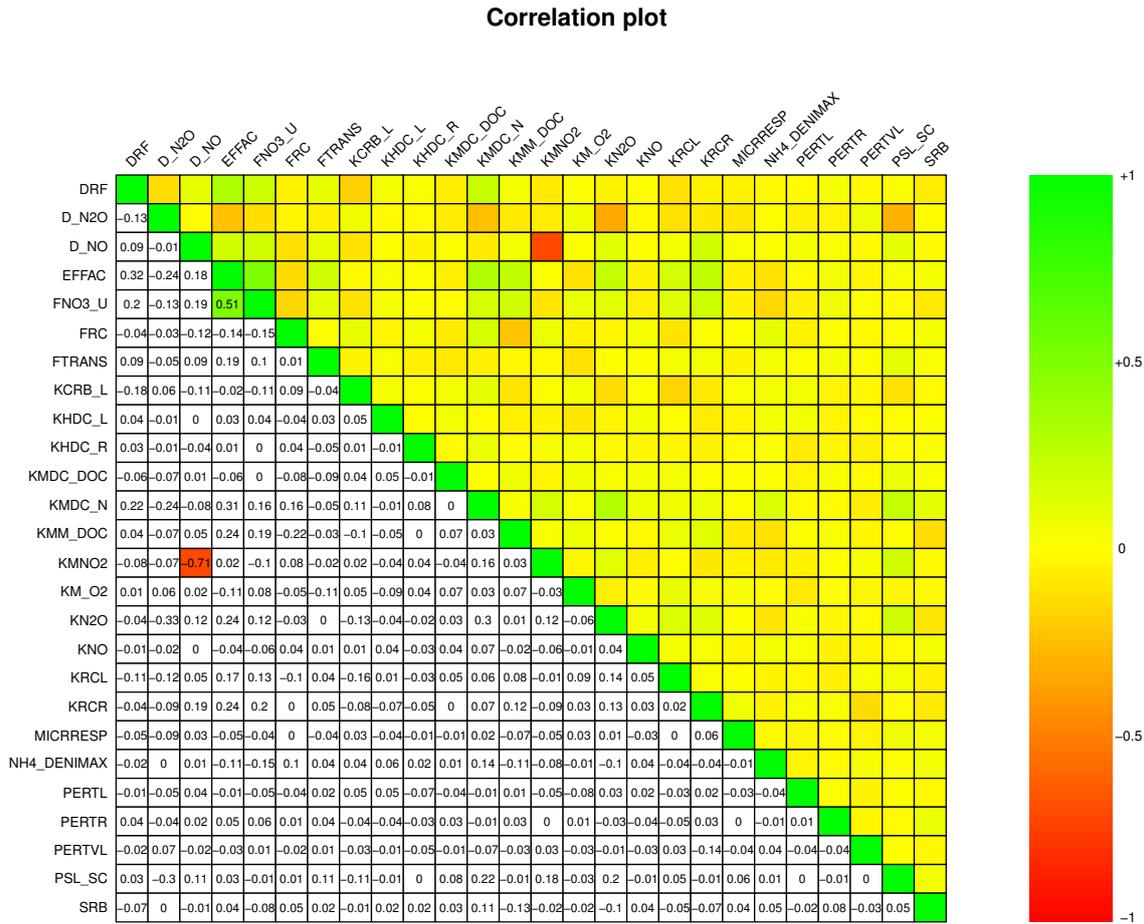


Fig. 8. Correlations of all post burn-in parameter samples.

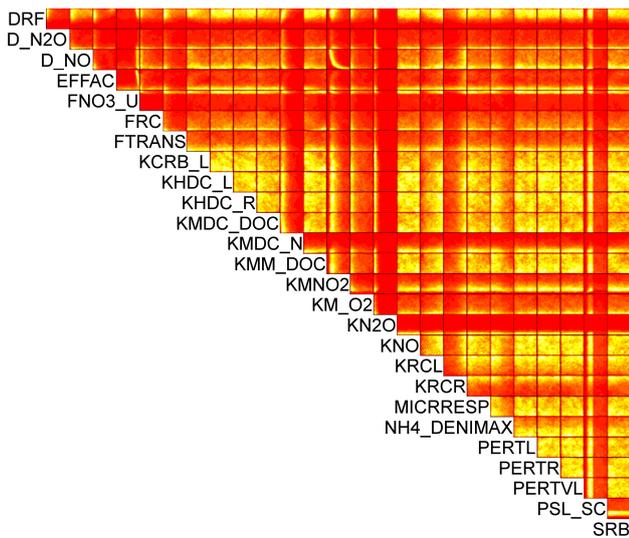


Fig. 9. Heat-maps of pair-wise marginal distributions; brighter polygons show higher posterior values.

The parameter distribution was estimated while running the model using the default parameter values for the remaining sub-modules (e.g., watercycle, plant growth). Hence, our estimation neglects dependencies between parameters of the soilchemistry submodule and parameters of other submodules. The posterior distribution is, thus, only estimated given the default values of the remaining modules. Changing these parameters and, therefore, the behaviour of the complex system will most likely change the distribution of the estimated parameters. Nevertheless, due to the complexity of LandscapeDNDC including parameters of other modules would complicate and lengthen the approach tremendously. More additional parameters would increase the degrees of freedom and, thus, the number of iterations and time to fully explore the parameter space. Even though we focused only on a subset of all model parameters our results show that parameter uncertainty could only be reduced for a subset of selected parameters. This may indicate model over-parameterisation and Bayesian calibration can give valuable guidance which parameters and module functions need to be focused on to allow further model improvement.

**Table 2.** Summary of marginal posterior parameter distribution and prior ranges (column 2). Posterior SD and skewness were estimated, whereas the prior SD was analytically calculated.

Parameter	Range <sub>prior</sub>	$\theta_{\text{MAP}}$	95 % cred. interval	Description	SD <sub>prior</sub>	SD <sub>post</sub>	$\frac{\text{SD}_{\text{post}}}{\text{SD}_{\text{prior}}}$	Skewness
D_N2O	[0.001, 0.150]	3.34e−03	[0.002, 0.144]	right-skew.	0.043	0.045	1.00	0.636
D_NO	[0.010, 0.150]	4.84e−02	[0.013, 0.145]	right-skew.	0.040	0.041	1.00	0.326
DRF	[5e−04, 0.055]	5.49e−02	[0.022, 0.055]	left-skew.	0.016	0.009	0.58	−0.564
EFFAC	[0.282, 0.955]	8.31e−01	[0.287, 0.910]	right-skew.	0.192	0.200	1.00	0.611
FNO3_U	[0.375, 1.000]	9.23e−01	[0.405, 0.995]	left-skew.	0.180	0.186	1.00	−0.846
FRC	[0.010, 0.400]	2.74e−02	[0.013, 0.385]	right-skew.	0.113	0.115	1.00	0.348
FTRANS	[0.000, 0.050]	3.53e−02	[4.8e−04, 0.048]	right-skew.	0.014	0.015	1.00	0.526
KCRB_L	[1.500, 4.000]	3.22e+00	[1.553, 3.932]	flat	0.722	0.726	1.00	0.122
KHDC_L	[0.001, 0.030]	2.89e−02	[0.002, 0.029]	flat	0.008	0.008	0.99	−0.033
KHDC_R	[0.001, 0.015]	2.20e−03	[0.001, 0.015]	flat	0.004	0.004	1.00	−0.061
KM_O2	[0.100, 0.990]	1.13e−01	[0.107, 0.950]	right-skew.	0.257	0.261	1.00	0.500
KMDC_DOC	[5e−04, 0.025]	8.25e−04	[0.001, 0.025]	flat	0.007	0.007	1.00	0.114
KMDC_N	[0.004, 0.245]	5.53e−02	[0.008, 0.214]	right-skew.	0.070	0.055	0.79	1.217
KMM_DOC	[1e−04, 0.009]	8.23e−03	[3.2e−04, 0.009]	flat	0.003	0.003	1.00	0.105
KMNO2	[0.002, 0.075]	4.13e−02	[0.014, 0.070]	right-skew.	0.021	0.016	0.74	0.834
KN2O	[5e−04, 0.090]	1.01e−02	[0.001, 0.019]	right-skew.	0.026	0.005	0.18	2.312
KNO	[5e−04, 0.025]	9.53e−03	[0.001, 0.024]	flat	0.007	0.007	1.00	−0.010
KRCL	[0.010, 0.900]	2.20e−01	[0.056, 0.886]	left-skew.	0.257	0.246	0.95	−0.289
KRCR	[0.050, 0.300]	2.65e−01	[0.061, 0.297]	left-skew.	0.072	0.071	0.98	−0.568
MICRRESP	[0.040, 0.120]	5.06e−02	[0.042, 0.118]	flat	0.023	0.023	1.00	0.116
NH4_DENIMAX	[0.700, 0.980]	8.21e−01	[0.705, 0.967]	right-skew.	0.081	0.078	0.96	0.342
PERTL	[2e−04, 7e−04]	6.38e−04	[2.6e−04, 7.4e−04]	flat	1.4e−04	1.4e−4	0.99	0.008
PERTR	[4e−05, 2e−04]	8.53e−05	[4.4e−05, 1.9e−04]	flat	4.6e−05	4.6e−5	1.00	0.038
PERTVL	[5e−04, 0.015]	9.16e−03	[0.001, 0.015]	flat	0.004	0.004	0.99	−0.024
PSL_SC	[0.003, 0.030]	1.21e−02	[0.004, 0.028]	right-skew.	0.008	0.007	0.82	0.866
SRB	[0.500, 0.990]	5.43e−01	[0.512, 0.977]	flat	0.141	0.142	1.00	0.058

By simultaneously calibrating soil N<sub>2</sub>O, NO and CO<sub>2</sub> emissions, we use a multi-objective (here three objectives) framework, so that e.g., a worsening of CO<sub>2</sub> estimation can be compensated by an improvement in NO or N<sub>2</sub>O estimation. Gathering additional data (e.g., from different forests sites) may help to reduce uncertainty for these parameters. However, multiple parameter solutions do not affect the process of uncertainty estimation of soil-atmosphere gas fluxes modelled by LandscapeDNDC, as the posterior parameter solution is used (including all parameter constellations) to generate the distribution of simulated emissions.

The large number of parameters chosen, the complexity of the LandscapeDNDC model (simulating the entire C, N and water fluxes of terrestrial ecosystems), as well as a narrow shaped posterior distribution as a result of a detailed dataset (Arhonditsis et al., 2008; Rahn et al., 2011; Clark, 2005; van Oijen et al., 2011), reduces the acceptance-rate. Consequently, slow convergence rates of the chains were observed. While estimating parameter EFFAC (partitioning of CO<sub>2</sub> and DOC production during microbial decomposition of organic matter) two of the chains at first showed a bimodal shape, the remaining chains sampled distinctly different modes. Therefore, the number of required iterations to reach convergence was substantially higher compared to the other parameters.

Hence, it took 165 000 iterations per chain, which required in total approximately four months computation time. In particular for the parameter EFFAC running four independent chains in parallel, demonstrated to be a more reliable and also necessary procedure to guarantee proper sampling from the posterior distribution.

The knowledge of all complex parameter dependencies helps to understand and improve the reliability of future model simulations and additionally to quantify the uncertainty of the simulated gas fluxes (N<sub>2</sub>O, NO, CO<sub>2</sub>) associated with model parameter uncertainty. As we use samples from the joint posterior distribution, we achieve more reliable uncertainty approximations of soil GHG exchange than by simply using samples of each marginal parameter distribution.

As we simultaneously calibrated the model parameters with data for three soil trace gas fluxes (N<sub>2</sub>O, NO and CO<sub>2</sub>) spanning six observation years, the parameter calibration results are a compromise for all years and the respective gas fluxes. Hence, better model simulation results are very likely to be obtained if single years or only one out of the three trace gases would have been chosen. Since the model is just an expert representation of the “real world” one cannot expect that simulation results and flux observations for all years and all

**Table 3.** Summary of cumulated measured and simulated emissions of NO, N<sub>2</sub>O and CO<sub>2</sub>. Simulated fluxes were only cumulated if corresponding periods with observations were available. Values in Brackets are calculated after freeze-thaw events.

Soil flux		1994	1995	1996	1997	2002	2003	Total	2004	2005	2006	2007	Total	
NO [kg N ha <sup>-1</sup> ]	No. of days	357	341	350	359	275	208	1890	162	322	263	263	1010	
	Minimum	3.86	3.98	3.60	3.62	2.95	2.24		1.43	3.19	2.15	2.70		
	Q <sub>0.005</sub>	6.10	5.48	4.92	5.85	4.61	3.49		2.11	4.72	3.57	4.35		
	Mean	8.14	7.25	6.65	7.98	6.09	4.91		3.14	6.43	5.27	5.82		
	St. dev	0.95	0.85	0.85	0.97	0.69	0.62		0.52	0.83	0.78	0.71		
	Q <sub>0.995</sub>	10.45	9.40	8.84	10.28	7.81	6.37		4.45	8.50	7.07	7.57		
	Maximum	11.57	10.46	9.98	11.46	8.63	7.76		5.12	9.42	7.92	8.29		
	Best	7.05	6.42	5.91	6.75	5.50	4.23		35.85	2.46	5.49	4.25	4.95	17.16
	Measured	6.23	8.16	8.69	6.98	4.24	6.73		41.03	3.62	5.46	8.64	4.38	22.11
N <sub>2</sub> O [kg N ha <sup>-1</sup> ]	No. of days	345	358	343	346	343	340	2075	296	343	264	294	1197	
	Minimum	0.32	0.29	0.25	0.29	0.29	0.23		0.20	0.27	0.23	0.22		
	Q <sub>0.005</sub>	0.40	0.39	0.32	0.36	0.36	0.30		0.26	0.33	0.29	0.28		
	Mean	0.82	0.76	0.66	0.75	0.72	0.73		0.63	0.67	0.63	0.59		
	St. dev	0.31	0.25	0.22	0.28	0.24	0.33		0.27	0.23	0.25	0.22		
	Q <sub>0.995</sub>	1.69	1.42	1.23	1.51	1.35	1.64		1.37	1.29	1.34	1.20		
	Maximum	2.23	1.78	1.60	1.82	1.64	1.97		1.66	1.58	1.69	1.49		
	Best	0.51	0.55(0.51)	0.55(0.38)	0.48(0.40)	0.54	0.41(0.34)		3.02(2.68)	0.37	0.45(0.37)	0.39(0.3)	0.39	1.60(1.43)
	Measured	0.39	0.80(0.75)	2.90(0.89)	0.61(0.25)	0.65	0.36(0.21)		5.72(3.29)	0.16	0.97(0.74)	2.14(0.51)	0.47	3.74(1.88)
CO <sub>2</sub> [kg C ha <sup>-1</sup> ]	No. of days		287	355	362			1004	299	334	331	228	1192	
	Minimum		5019	4212	5782				4187	4925	4916	3442		
	Q <sub>0.005</sub>		6135	5263	6982				4984	6001	5923	4162		
	Mean		8073	7163	9166				6487	7937	7776	5504		
	St. dev		765	756	873				610	771	736	537		
	Q <sub>0.995</sub>		10048	9142	11452				8104	9954	9701	6917		
	Maximum		10660	9761	12174				8606	10584	10323	7364		
	Best		8133	7250	9282				24665	6570	8036	7852	5576	28035
	Measured		10673	8813	7740				27226	5294	7332	7556	3913	24095

**Table 4.** Root-mean-squared error (RMSE) per year and soil-atmosphere gas-flux for the best simulation (RMSE( $\theta_{MAP}$ )) and the distribution of the gas-flux simulations. The minimal distance to the 99 % credible intervals was used to calculate the RMSE of the distribution (RMSE(CI)). Values in brackets are calculated using simulated emissions after freeze-thaw events.

Soil flux		1994	1995	1996	1997	2002	2003	2004	2005	2006	2007
NO [g N ha <sup>-1</sup> d <sup>-1</sup> ]	No of days	357	341	350	359	275	208	162	322	263	263
	RMSE(CI)	2.40	4.54	9.42	2.86	3.20	15.27	6.44	8.26	18.95	5.34
	RMSE( $\theta_{MAP}$ )	6.66	11.68	15.92	9.20	7.94	22.66	12.46	14.45	28.32	11.57
N <sub>2</sub> O [g N ha <sup>-1</sup> d <sup>-1</sup> ]	No of days	345	358	343	346	343	340	296	343	264	294
	RMSE(CI)	0.21	0.86(0.83)	16.10(7.92)	2.56(0.42)	0.49	0.69(0.31)	0.40	3.47(0.50)	21.12(2.60)	0.14
	RMSE( $\theta_{MAP}$ )	0.53	1.82(1.77)	17.01(8.22)	3.12(0.77)	1.45	1.09(0.67)	0.83	4.19(2.17)	21.80(3.10)	1.06
CO <sub>2</sub> [g C ha <sup>-1</sup> d <sup>-1</sup> ]	No of days		287	355	362			299	334	331	228
	RMSE(CI)		9718	5977	5373			8571	9846	11849	11949
	RMSE( $\theta_{MAP}$ )		15344	9986	8989			11623	13106	15806	16446

gases are in perfect agreement. However, the results show that the LandscapeDNDC model is able to follow most of the dynamics as observed in field measurements and to approximate annual total emissions (see Table 3) with a certain accuracy (RMSE NO: 2.40 to 18.95 g N ha<sup>-1</sup> d<sup>-1</sup>, N<sub>2</sub>O: 0.21 to 21.12 g N ha<sup>-1</sup> d<sup>-1</sup>, CO<sub>2</sub>: 5.4 to 11.9 kg C ha<sup>-1</sup> d<sup>-1</sup>, Table 4) not only for the years which were used for model calibration but also for independent observation years.

Lowest agreement between measured and simulated fluxes was obtained for N<sub>2</sub>O. Most of the discrepancy is due to the inability of the current LandscapeDNDC version to simulate

freeze-thaw N<sub>2</sub>O pulse emission events. Since up to now no frost-thaw process descriptions were implemented into LandscapeDNDC, the calibration procedure was not able to fit the model to these fluxes sufficiently. At the Höglwald spruce site as well as in other temperate ecosystems exposed to severe winter freezing periods, freeze-thaw N<sub>2</sub>O fluxes may dominate annual N<sub>2</sub>O fluxes (Papen and Butterbach-Bahl, 1999; Wolf et al., 2010), so that a failure to simulate N<sub>2</sub>O fluxes during freeze-thaw periods must lead to incorrect annual flux estimates. However, the comparison of N<sub>2</sub>O data excluding freeze-thaw periods reveal, that simulations

of N<sub>2</sub>O fluxes by LandscapeDNDC are generally in the same range as the measurements for both the calibration and the validation set. Nevertheless, due to the importance of freeze-thaw emissions for the annual N<sub>2</sub>O budget there is an urgent need to further develop and implement model algorithms describing underlying processes of freeze-thaw based N<sub>2</sub>O production and emission in/from soils (e.g. Wolf et al., 2011).

Also with regard to soil NO and CO<sub>2</sub> fluxes we identified short-comes of the applied LandscapeDNDC. For example, higher NO emissions in the summer period in 1995 and 1996 were systematically underestimated, while soil CO<sub>2</sub> emissions tended to be overestimated in the end of spring and beginning of summer and underestimated in subsequent summer days. This points either towards insufficient process descriptions, which have already been suggested earlier (Stange et al., 2000), or to problems with model initialisation. We limited the calibration procedure to a subset of 26 most influential parameters describing C and N turnover and production, consumption and emission processes of N<sub>2</sub>O, NO and CO<sub>2</sub> in soils. To allow a more time efficient calibration, we excluded parameters describing soil water and vegetation dynamics. Nevertheless, the above-mentioned failures to accurately describe soil NO and CO<sub>2</sub> fluxes for all seasons point towards the necessity to recheck simulated soil water and vegetation dynamics in LandscapeDNDC.

However, in total the measurements of the calibration and the validation set were covered reasonably well, in particular if we consider that not all sources of errors (i.e., structural model error, input data error) were included. In order to achieve improved approximations of the uncertainty of N<sub>2</sub>O, NO and CO<sub>2</sub> emissions, a stochastic error term could be included in future research, e.g., by setting up a hierarchical Bayesian framework, to account for model misspecifications (Rahn et al., 2011; Arhonditsis et al., 2008).

## 7 Conclusions

Following the identification of the 26 most sensitive parameters out of the total set of 67 parameters used in the submodule soil-chemistry, describing soil emission of N<sub>2</sub>O, NO and CO<sub>2</sub> in the biogeochemical model LandscapeDNDC, we successfully implemented a Bayesian calibration to estimate the joint posterior distribution of the most influential model parameters. To ensure that the posterior distribution of parameters was assessed, we used a multi-chain approach and tested for convergence of the Markov chain by the objective criteria developed by Gelman et al. (2003). In contrast to the a priori assumption of a uniform distribution of parameter values over a given range the posterior parameter distribution showed a more distinct pattern, including all complex parameter dependencies. Bayesian calibration reduced the prior uncertainty (by up to 82 %) of 15 out of 26 parameters. This knowledge of the posterior probability distribution is of outstanding importance to guide future model development, e.g.,

to inform experimentalists which parameters need attention for further investigation.

The comparison of simulated soil N<sub>2</sub>O, NO and CO<sub>2</sub> emissions to measured flux data over the six observation years used in the calibration process showed high agreement. The same is true for independent validation data, including observations of four other years. Hence, we were able to quantify the parameter-induced uncertainty of the total simulated N<sub>2</sub>O, NO and CO<sub>2</sub> emission. However, further studies need to consider other uncertainties such as a model error in order to estimate the total uncertainty of simulated soil fluxes of N<sub>2</sub>O, NO and CO<sub>2</sub>.

In our study, freeze-thaw events could not be reproduced, as underlying processes are not included in the LandscapeDNDC version used in this study. Since these events can potentially have a strong impact on the total annual soil N<sub>2</sub>O emission, future model development and implementation of freeze-thaw algorithms is intended.

*Acknowledgements.* This work was part of the EU project NitroEurope, funded under the 6th EC Framework Programme for Research and Technological Development. This project received additional funding from the German Science Foundation (DFG) under contract no. BU1173/12-1. C. Werner would like to acknowledge financial support by the research funding programme LOEWE “Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz” of Hesse’s Ministry of Higher Education, Research, and the Arts.

Edited by: U. Skiba

## References

- Arhonditsis, G. B., Perhar, G., Zhang, W., Massos, E., Shi, M., and Das, A.: Addressing equifinality and uncertainty in eutrophication models, *Water Resour. Res.*, 44, W01420, doi:10.1029/2007WR005862, 2008.
- Butterbach-Bahl, K., Stange, F., Papen, H., and Li, C.: Regional inventory of nitric oxide and nitrous oxide emissions for forest soils of southeast Germany using the biogeochemical model PnET-N-DNDC, *J. Geophys. Res.-Atmos.*, 106, 34155–34166, doi:10.1029/2000JD000173, 2001.
- Butterbach-Bahl, K., Rothe, A., and Papen, H.: Effect of tree distance on N<sub>2</sub>O and CH<sub>4</sub>-fluxes from soils in temperate forest ecosystems, *Plant Soil*, 240, 91–103, doi:10.1023/A:1015828701885, 2002.
- Clark, J.: Why environmental scientists are becoming Bayesians, *Ecol. Lett.*, 8, 2–14, doi:10.1111/j.1461-0248.2004.00702.x, 2005.
- Del Grosso, S., Parton, W., Mosier, A., Walsh, M., Ojima, D., and Thornton, P.: DAYCENT national-scale simulations of nitrous oxide emissions from cropped soils in the United States, *J. Environ. Qual.*, 35, 1451–1460, 2006.
- Ellison, A.: An introduction to Bayesian inference for ecological research and environmental decision-making, *Ecol. Appl.*, 6, 1036–1046, doi:10.2307/2269588, 1996.

- Gasche, R. and Papen, H.: A 3-year continuous record of nitrogen trace gas fluxes from untreated and limed soil of a N-saturated spruce and beech forest ecosystem in Germany – 2. NO and NO<sub>2</sub> fluxes, *J. Geophys. Res.-Atmos.*, 104, 18505–18520, doi:10.1029/1999JD900294, 1999.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D.: *Bayesian Data Analysis*, Chapman and Hall, 2 Edn., London, UK, 2003.
- Gilks, W., Richardson, S., and Spiegelhalter, D.: *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, UK, 1996.
- Grote, R.: Sensitivity of volatile monoterpene emission to changes in canopy structure: a model-based exercise with a process-based emission model, *New Phytologist*, 173, 550–561, doi:10.1111/j.1469-8137.2006.01946.x, 2007.
- Grote, R., Lehmann, E., Brümmer, C., Brüggemann, N., Szarzynski, J., and Kunstmann, H.: Modelling and observation of biosphere-atmosphere interactions in natural savannah in Burkina Faso, West Africa, *Phys. Chem. Earth.*, 34, 251–260, doi:10.1016/j.pce.2008.05.003, 2009.
- Grote, R., Kiese, R., Grünwald, T., Ourcival, J.-M., and Granier, A.: Modelling forest carbon balances considering tree mortality and removal, *Agr. Forest Meteorol.*, 151, 179–190, doi:10.1016/j.agrformet.2010.10.002, 2011.
- Haas, E., Klatt, S., Fröhlich, A., Kraft, P., Werner, C., Kiese, R., Grote, R., Breuer, L., and Butterbach-Bahl, K.: LandscapeDNDC: A process model for simulation of biosphere-atmosphere-hydrosphere exchange processes at site and regional scale, *Landscape Ecol.*, 1–22, doi:10.1007/s10980-012-9772-x, 2012.
- Hamby, D. M.: A review of techniques for parameter sensitivity analysis of environmental models, *Environ. Monit. Assess.*, 32, 135–154, doi:10.1007/BF00547132, 1994.
- Kass, R., Carlin, B., Gelman, A., and Neal, R.: Markov chain Monte Carlo in practice: a roundtable discussion, *Am. Stat.*, 52, 93–100, doi:10.2307/2685466, 1998.
- Kesik, M., Ambus, P., Baritz, R., Brüggemann, N., Butterbach-Bahl, K., Damm, M., Duyzer, J., Horváth, L., Kiese, R., Kitzler, B., Leip, A., Li, C., Pihlatie, M., Pilegaard, K., Seufert, S., Simpson, D., Skiba, U., Smiatek, G., Vesala, T., and Zechmeister-Boltenstern, S.: Inventories of N<sub>2</sub>O and NO emissions from European forest soils, *Biogeosciences*, 2, 353–375, doi:10.5194/bg-2-353-2005, 2005.
- Kesik, M., Brüggemann, N., Forkel, R., Kiese, R., Knoche, R., Li, C., Seufert, G., Simpson, D., and Butterbach-Bahl, K.: Future scenarios of N<sub>2</sub>O and NO emissions from European forest soils, *J. Geophys. Res.*, 111, 148–227, doi:10.1029/2005JG000115, 2006.
- Kiese, R., Li, C., Hilbert, D., Papen, H., and Butterbach-Bahl, K.: Regional application of PnET-N-DNDC for estimating the N<sub>2</sub>O source strength of tropical rainforests in the Wet Tropics of Australia, *Glob. Change Biol.*, 11, 128–144, doi:10.1111/j.1365-2486.2004.00873.x, 2005.
- Kiese, R., Heinzeller, C., Werner, C., Wochele, S., Grote, R., and Butterbach-Bahl, K.: Quantification of nitrate leaching from German forest ecosystems by use of a process oriented biogeochemical model, *Environ. Pollut.*, 159, 3204–3214, doi:10.1016/j.envpol.2011.05.004, 2011.
- Klemetsson, L., Jansson, P., Gustafsson, D., Karlberg, L., Weslien, P., von Arnold, K., Ernfors, M., Langvall, O., and Lindroth, A.: Bayesian calibration method used to elucidate carbon turnover in forest on drained organic soil, *Biogeochemistry*, 89, 61–79, 2008.
- Lehuger, S., Gabrielle, B., van Oijen, M., Makowski, D., Geron, J., Morvan, T., and Hénault, C.: Bayesian-calibration of the nitrous oxide emission module of an agro-ecosystem model, *Agr. Ecosyst. Environ.*, 133, 208–222, 2009.
- Li, C., Frolking, S., and Frolking, T. A.: A model of nitrous-oxide evolution from soil driven by rainfall events. 1. Model structure and sensitivity, *J. Geophys. Res.*, 97, 9759–9776, 1992.
- Li, C., Aber, J., Stange, F., Butterbach-Bahl, K., and Papen, H.: A process-oriented model of N<sub>2</sub>O and NO emissions from forest soils: 1. model development, *J. Geophys. Res.-Atmos.*, 105, 4369–4384, doi:10.1029/1999JD900949, 2000.
- Li, C., Mosier, A., Wassmann, R., Cai, Z., Zheng, X., Huang, Y., Tsuruta, H., Boonjawat, J., and Lantin, R.: Modeling greenhouse gas emissions from rice-based production systems: Sensitivity and upscaling, *Global Biogeochem. Cy.*, 18, GB1043, doi:10.1029/2003GB002045, 2004.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087–1092, 1953.
- Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174, available online: <http://www.jstor.org/stable/1269043>, 1991.
- Papen, H. and Butterbach-Bahl, K.: A 3-year continuous record of nitrogen trace gas fluxes from untreated and limed soil of a N-saturated spruce and beech forest ecosystem in Germany – 1. N<sub>2</sub>O emissions, *J. Geophys. Res.-Atmos.*, 104, 18487–18503, doi:10.1029/1999JD900293, 1999.
- Pathak, H., Li, C., and Wassmann, R.: Greenhouse gas emissions from Indian rice fields: calibration and upscaling using the DNDC model, *Biogeosciences*, 2, 113–123, doi:10.5194/bg-2-113-2005, 2005.
- Potter, C., Matson, P., Vitousek, P., and Davidson, E.: Process modeling of controls on nitrogen trace gas emissions from soils worldwide, *J. Geophys. Res.-Atmos.*, 101, 1361–1377, doi:10.1029/95JD02028, 1996.
- Rahn, K.-H., Butterbach-Bahl, K., and Werner, C.: Selection of likelihood parameters for complex models determines the effectiveness of Bayesian calibration, *Ecol. Inform.*, doi:10.1016/j.ecoinf.2011.08.002, 2011.
- Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P.: A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29, 1586–1597, doi:10.1016/j.advwatres.2005.11.013, available online: <http://www.sciencedirect.com/science/article/pii/S03091708050%02903>, 2006.
- Reinds, G. J., van Oijen, M., Heuvelink, G. B., and Kros, H.: Bayesian calibration of the VSD soil acidification model using European forest monitoring data, *Geoderma*, 146, 475–488, 2008.
- Salas, W., Boles, S., Li, C., Yeluripati, J. B., Xiao, X., Frolking, S., and Green, P.: Mapping and modelling of greenhouse gas emissions from rice paddies with satellite radar observations and the DNDC biogeochemical model, *Aquat. Conserv.*, 17, 319–329, doi:10.1002/aqc.837, 2007.
- Saltelli, A. (Ed.): *Global Sensitivity Analysis: the Primer*, Wiley, Chichester, West Sussex, 2008.
- Stange, F., Butterbach-Bahl, K., Papen, H., Zechmeister-Boltenstern, S., Li, C., and Aber, J.: A process-oriented

- model of N<sub>2</sub>O and NO emissions from forest soils. 2. sensitivity analysis and validation, *J. Geophys. Res.-Atmos.*, 105, 4385–4398, doi:10.1029/1999JD900948, 2000.
- Svensson, M., Jansson, P.-E., Gustafsson, D., Kleja, D. B., Langvall, O., and Lindroth, A.: Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand, *Ecol. Model.*, 213, 331–344, 2008.
- Toft, N., Innocent, G. T., Gettinby, G., and Reid, S. W.: Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard, *Prev. Vet. Med.*, 79, 244–256, doi:10.1016/j.prevetmed.2007.01.003, 2007.
- van Oijen, M., Rougier, J., and Smith, R.: Bayesian calibration of process-based forest models: bridging the gap between models and data, *Tree Physiol.*, 25, 915–927, 2005.
- van Oijen, M., Cameron, D., Butterbach-Bahl, K., Farahbakhsh, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., and Yeluripati, J.: A Bayesian framework for model calibration, comparison and analysis: application to four models for the biogeochemistry of a Norway spruce forest, *Agr. Forest Meteorol.*, doi:10.1016/j.agrformet.2011.06.017, 2011.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, doi:10.1029/2002WR001642, 2003.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720, 2008.
- Werner, C., Zheng, X., Tang, J., Xie, B., Liu, C., Kiese, R., and Butterbach-Bahl, K.: N<sub>2</sub>O, CH<sub>4</sub> and CO<sub>2</sub> emissions from seasonal tropical rainforests and a rubber plantation in Southwest China, *Plant Soil*, 289, 335–353, doi:10.1007/s11104-006-9143-y, 2006.
- Werner, C., Butterbach-Bahl, K., Haas, E., Hickler, T., and Kiese, R.: A global inventory of N<sub>2</sub>O emissions from tropical rainforest soils using a detailed biogeochemical model, *Global Biogeochem. Cy.*, 21, GB3010, doi:10.1029/2006GB002909, 2007.
- Werner, C., Haas, E., Grote, R., Gauder, M., Graeff-Hönninger, S., Claupein, W., and Butterbach-Bahl, K.: Biomass production potential from *Populus* short rotation systems in Romania, *Global Change Biol.*, 4, 642–653, doi:10.1111/j.1757-1707.2012.01180.x, 2012.
- Wikle, C.: Hierarchical Bayesian models for predicting the spread of ecological processes, *Ecology*, 84, 1382–1394, doi:10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2, 2003.
- Winiwarter, W. and Rypdal, K.: Assessing the uncertainty associated with national greenhouse gas emission inventories: a case study for Austria, *Atmos. Environ.*, 35, 5425–5440, doi:10.1016/S1352-2310(01)00171-6, available online: <http://www.sciencedirect.com/science/article/pii/S1352231001001716>, 2001.
- Wolf, B., Zheng, X., Brüggemann, N., Chen, W., Dannenmann, M., Han, X., Sutton, M. A., Wu, H., Yao, Z., and Butterbach-Bahl, K.: Grazing-induced reduction of natural nitrous oxide release from continental steppe, *Nature*, 464, 881–884, doi:10.1038/nature08931, 2010.
- Wolf, B., Chen, W., Brüggemann, N., Zheng, X., Pumpanen, J., and Butterbach-Bahl, K.: Applicability of the soil gradient method for estimating soil-atmosphere CO<sub>2</sub>, CH<sub>4</sub> and N<sub>2</sub>O fluxes for steppe soils in Inner Mongolia, *J. Plant Nutr. Soil Sc.*, 174, 359–372, doi:10.1002/jpln.201000150, 2011.
- Wu, X., Brüggemann, N., Gasche, R., Shen, Z., Wolf, B., and Butterbach-Bahl, K.: Environmental controls over soil-atmosphere exchange of N<sub>2</sub>O, NO, and CO<sub>2</sub> in a temperate Norway spruce forest, *Global Biogeochem. Cy.*, 24, GB2012, doi:10.1029/2009GB003616, 2010.