Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model

Simon Lehuger, Benoit Gabrielle, M. van Oijen, David Makowski, Jean Claude Germon, Thierry Morvan, Catherine Hénault

Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model

S. Lehuger[a 1], B. Gabrielle[a], M. Van Oijen[b], D. Makowski[c], J.-C. Germon[d], T. Morvan[e], C. Hénault[d]

a: Institut National de la Recherche Agronomique, UMR 1091 INRA-AgroParisTech Environnement et Grandes Cultures, 78850 Thiverval-Grignon, France
b: Centre for Ecology and Hydrology-Edinburgh, Bush Estate, Penicuik, UK
c: Institut National de la Recherche Agronomique, UMR 211 INRA-AgroParisTech Agronomie, Thiverval-Grignon, France
d: Institut National de la Recherche Agronomique, UMR 1229 Microbiologie du sol et de l'environnement, Dijon, France
e: Institut National de la Recherche Agronomique, UMR 1069 INRA-Agrocampus Sol Agro et hydrosystème Spatialisation, Rennes, France

[1]Corresponding author: UMR1091 INRA, AgroParisTech Environnement et Grandes Cultures, 78850 Thiverval-Grignon, France. E-mail: Simon.Lehuger@grignon.inra.fr Fax: (+33) 1 30 81 55 63 Phone: (+33) 1 30 81 55 24

Abstract

Nitrous oxide (N_2O) is the main biogenic greenhouse gas contributing to the global warming potential (GWP) of agro-ecosystems. Evaluating the impact of agriculture on climate therefore requires a capacity to predict the net exchanges of this gas in relation to environmental conditions and crop management. The biophysical crop model CERES-EGC is designed to predict the productivity and GWP of agro-ecosystems by simulating C and N dynamics, including N_2O emissions from soils, on a daily time step, as driven by the nitrification and denitrification pathways. These two microbiological processes are modelled as the product of a potential rate with three dimensionless factors related to soil water content, nitrogen content and temperature, of which a fixed site-specific proportion is evolved as N_2O. These equations form the N_2O submodel of CERES-EGC, and involve a total set of 15 parameters. Four of those are site-specific and should be measured on site, while the other 11 are considered global, i.e. fixed over time and space. Accurate estimates of the global parameters should be sought prior to extrapolating the model to make predictions in new situations. Here, we used Bayesian calibration to that purpose using a database of N_2O flux measurements including seven different field-sites in France. First, we gathered prior information on the model parameters based on literature review, and assigned them uniform probability distributions. A Bayesian method based on the Metropolis-Hastings algorithm was subsequently used to update the parameter distributions for each field site. Three parallel Markov chains were run to ensure a convergence of the algorithm. This site-specific calibration significantly reduces the model prediction error across the field sites, along with its overall uncertainty, compared to the initial parameter setting. The root mean square error (RMSE) of predictions computed with posterior parameters values was thus reduced by 73% on average in comparison with the prior estimates. The RMSE declined from 39 to 6 g N_2O-N ha^{-1} day^{-1} on average. The Bayesian calibration was also applied to all the data sets simultaneously, to obtain

better global estimates for the parameters initially deemed universal. This made it possible to reduce the RMSE by 33% on average, compared to the uncalibrated model. These global parameter values may be used to obtain more realistic estimates of N_2O emissions from arable soils at regional or continental scales.

Keywords

Bayesian calibration, Parameter uncertainty, CERES-EGC, Nitrous oxide, Markov Chain Monte Carlo, NitroEurope

1 Introduction

While food supply for increasing population is becoming one of the alarming question worldwide, we are faced with the growing environmental footprint of agriculture due to land use change and management intensification (Kiers et al., 2008). Assessing the contribution of agriculture to climate change is one of the key question addressed to environmental scientists who should help to identify measures to reduce the burden of agriculture in global warming. Soils are the main source of nitrous oxide (N_2O) in the atmosphere due to microbial processes of nitrification and denitrification. By intensively using N-fertilisers, agriculture amplifies these two processes and hence, agro-ecosystems contribute 55-65% of the global anthropogenic emissions of N_2O and are the most responsible for the increase of N_2O atmospheric concentration compared to other ecosystems or activity sectors (Smith et al., 2007). The use of agro-ecosystem models facilitates predictions of N_2O emissions from arable soils at the plot scale and offers the unique mean to upscale the predictions at regional and continental scales (Butterbach-Bahl et al., 2004). Predictions of process-based models such as agro-ecosystem models are highly dependant on model parameters and uncertainty about their values inevitably induces uncertainty about model outputs. To facilitate decisions based on model, it requires first to estimate the parameter values and then to quantify the risk of error of prediction due to parameter estimates. Although model parameterisation and uncertainty analysis of process-based models are widely developed in the literature, they rarely are considered simultaneously. Bayesian calibration makes the combination of this two goals possible by providing estimates of parameters values under the form of probability density functions (pdfs) which are propagated to model outputs that can also be expressed as pdfs (Gallagher and Doherty, 2007). Probability density functions are initially the expression of current imprecise knowledge about model parameter values, this prior probability is then updated with the measured observations into posterior probability distribution by means

of Bayes' theorem.

In ecological and environmental sciences, Bayesian calibration has been applied for various different models and is actively developing for many types of models. For example, Hong et al. (2005) applied a Bayesian estimation to input parameters of a nitrogen cycle model that simulates N cycle at the watershed scale, Larssen et al. (2006) used a Bayesian approach for model calibration and uncertainty analysis of a hydrogeochemical model that simulates acidification effects on trout population health and Ricciuto et al. (2008) have developed a technique performing both calibration of all the parameters of a simple model of simulation of net ecosystem CO_2 exchanges and assimilation of hourly observations into the model. All these techniques were based on Markov Chain Monte Carlo (MCMC), a Bayesian technique that has demonstrated its superiority compared to other methods of parameter estimation. Qian et al. (2003) and Gallagher and Doherty (2007) demonstrated that MCMC methods are the most powerful methods compared to other Bayesian and frequentist methods of uncertainty analysis. In the same way, Makowski et al. (2002) demonstrated that the Metropolis-Hastings gives lower mean squared error of prediction than the Generalized Likelihood Uncertainty Estimation method (GLUE) in the case of parameter estimation of an agronomic model. The Bayesian methodology described by Van Oijen et al. (2005) has been applied to dynamic process-based forest models with the goal to calibrate model parameters with multiple observed data from forested experimental sites (Svensson et al., 2008; Klemedtsson et al., 2007). The technique is based on the Metropolis-Hastings algorithm that generates samples from high dimensional distributions under the form of Markov Chains Monte Carlo (MCMC) which approximate the posterior parameter distributions.

Although a large literature is developing about application of Bayesian techniques in environmental sciences, Bayesian approach has never been applied to process-based model of soil N_2O emission models. These models have been developed first to by-pass the expensive and time-consuming direct measurements of N_2O emissions on field and then to extrapolate emis-

sions over space and time. Indeed, models of N_2O emissions are definitely indispensable to facilitate simulation and interpretation of specific measuring sites, to make tests of different management and mitigation strategies possible for farmers and to carry out spatially-explicit inventories of N_2O emissions from agriculture. Predicting N_2O emissions from agro-ecosystems requires taking into account complex processes and interactions which originate from both environmental conditions and agricultural practises (Frolking et al., 1998). Several process-based models have been developed to simulate N_2O emissions from arable soils, including DAYCENT (Parton et al., 2001), DNDC (Li, 2000), FASSET (Chatskikh et al., 2005) and CERES-EGC (Gabrielle et al., 2006b). As recommended by IPCC, the tier 3 methodology allows the countries of Annex 1 to use these more complex models to quantify their N_2O budgets from agriculture in order to improve the accuracy of national inventories (IPCC, 2006). However these models have still high uncertainties coming from parameter values, driving variables and model structure (Gabrielle et al., 2006a).

The overall purpose of this study was to calibrate the parameters of the N_2O emissions module of the CERES-EGC agro-ecosystem model and to quantify uncertainty of model predictions by using a Bayesian calibration technique and a sophisticated procedure for analysis and diagnostics of MCMC chains. Measured observations were from seven field-sites in Northern France which represent major soil types, crops and crop management. Two additional specific objectives were to apply our Bayesian procedure both to the measured data sets successively and to the data sets simultaneously in order to find respectively, site-specific and universal parameter values with their uncertainty quantified. Our results will be useful in using the site-specific calibrated parameters to develop mitigation strategies in the measuring sites and the universal values of calibrated parameters to extrapolate the model at regional and continental scales. In addition, our study allows us to partly reach the objectives of the Plot Scale Modelling Component of the NitroEurope Integrated project (Sutton et al., 2007) in which we are actively involved and thus

to give a little advance to the question addressed by the project: "What is the effect of reactive nitrogen supply on the direction and magnitude of net greenhouse gas budgets for Europe?"

2 Material and Methods

We used a Bayesian calibration technique based on the Metropolis-Hastings algorithm to estimate the parameter probability density functions of the N_2O emissions module of the CERES-EGC model. The equations of the N_2O emissions module involve a total set of 15 parameters of which 11 were estimated by our procedure based on running three parallel Markov Chains Monte Carlo to ensure a convergence of the algorithm. We used a database of N_2O flux measurements including seven different field-sites in France in the goal of either to apply our procedure for each data sample successively and find the site-specific parameter estimates or to apply our calibration procedure to all the data samples simultaneously and find parameters estimates considered as universal.

2.1 The Bayesian approach

Bayesian methods are used to estimate model parameters by combining two sources of information: the prior information about parameter values and the measurements of output data. The prior information is based on expert knowledge, literature review or by measuring them directly on field or in laboratory and observations are generally direct-field measurements which assess the different fluxes between soil-crop-atmosphere compartments. Using the Bayes' theorem makes it possible to combine these two sources of information in order to calibrate the model parameters whereas most statistical methods only use the output data and does not provide probability distributions of the parameters. The uncertain parameters are random variables for which we assigned them a prior probability distribution. This probability distribution constitutes the prior uncertainty about parameter values and what we wanted was to reduce this uncertainty

by using the measured data. In our case, we specified lower and upper bounds of the parameters' uncertainty which define the prior parameter distributions as uniform. In addition, using this approach allowed us to analyse the uncertainty of model predictions by running the model with different parameter settings sampled from the calibrated parameter distributions and to quantify cross-correlations between parameters. Bayesian calibration generates the posterior parameter distribution which is given by Bayes' theorem:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \tag{1}$$

where we denote the parameters as θ and the vector of measurements as Y. $p(\theta)$ is the prior parameter distribution for θ, $p(\theta|Y)$ is the posterior parameter distribution, $p(Y)$ is a constant of proportionality that is not explicitly computed, and $p(Y|\theta)$ is the likelihood function for θ. The likelihood is the probability of the data Y given the parameters θ and is determined from the probability distribution of errors between observations and predictions. In our case, we assume that the errors are normally distributed with mean 0 and uncorrelated, and as probability densities may be very small numbers and to avoid rounding errors we assumed calculations using logarithms. The logarithm of the data likelihood is thus set up as follows:

$$logL = \sum_{j=1}^{K} \left(-0.5 \left(\frac{y_j - f(\omega_j;\theta)}{\sigma_i} \right)^2 - 0.5log(2\pi) - log(\sigma_i) \right) \tag{2}$$

where y_j is the j^{th} y value in the data set Y, ω_j is the vector of model input data associated with y_j, $f(\omega_j;\theta)$ is the model prediction of y_j, and K is the total number of observations in the data sets. The data sets Y are times-series of N_2O emission measurements and σ_i is the standard deviation of the N_2O measurements. We assumed that the model error can be attributed to additive measurement errors following Van Oijen et al. (2005) and in the same fashion as Svensson et al. (2008) and Klemedtsson et al. (2007).

2.2 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo (MCMC) technique that generates a sample of parameter vectors from the posterior distribution $p(Y|\theta)$ (Metropolis et al., 1953). First of all, the starting vector θ_0 of the Markov chain in the parameter space is chosen within the prior parameter space and then the next candidate vectors θ_i are generated for $i = 1, ..., N$ iterations as follows:

Step 1. Randomly generate a proposal parameter vector for a new candidate parameter vector

$$\theta^* = \theta_{i-1} + \delta \tag{3}$$

where δ is a random vector generated using a multivariate normal distribution;

Step 2. Calculate the ratio of the posterior probability of the new candidate over the posterior probability of the previous candidate:

$$\alpha = \frac{p(\theta^*|Y)}{p(\theta_{i-1}|Y)} = \frac{p(Y|\theta^*)p(\theta^*)}{p(Y|\theta_{i-1})p(\theta_{i-1})} \tag{4}$$

In our case, since calculations are made using logarithms, we compute the log of α as the difference between the log of the posterior probability of current point minus the log of posterior probability of the previous point.

Step 3. Accept θ^* if $log\alpha \geq u$ where u is an uniform random variable from an uniform distribution on the interval (0,1), else reject and $\theta_i = \theta_{i-1}$.

The new point θ^* is always accepted if its posterior value is no lower than the posterior value of θ_{i-1}. Once the chain has attained the N iterations, the chain must have converged to the target distribution which is the posterior parameter distribution $p(\theta|Y)$. Before running the algorithm, the chain length N (total number of iterations) and the proposal distribution δ for generating new candidates must have been pre-defined as well as the M number of first iterations that should be

discarded. The first iterations at the beginning of the chain need to be discarded while the MCMC "burns-in". In our case, we discard 10% of the total number of iterations at the beginning of the chain (Van Oijen et al., 2005). The chain length N is fixed between 10^4 and 10^5 iterations and depends on the convergence point of the Markov chains. The proposal distribution δ used for generating new candidate vectors is a Gaussian distribution with a mean of zero. In our case, we tuned the variance matrix Σ of δ so that the Markov chains explore the space of possible values for θ. We subsequently followed the procedure of Van Oijen et al. (2005) and we defined the variances equal to the square of 1 to 5 % of the prior parameter range ($\theta_{min} - \theta_{max}$) and zero covariances. The variances of Σ were tuned so that the fraction of accepted points was comprised between 20 to 30% during the test performed in the step 2 of the Metropolis-Hastings algorithm.

2.3 The CERES-EGC model

CERES-EGC was adapted from the CERES suite of soil-crop models (Jones and Kiniry, 1986), with a focus on the simulation of environmental outputs such as nitrate leaching, emissions of N_2O and nitrogen oxides (Gabrielle et al., 2006a). CERES-EGC runs on a daily time step, and requires daily rain, mean air temperature and Penman potential evapo-transpiration as forcing variables. The CERES models are available for a large number of crop species, which share the same soil components (Jones and Kiniry, 1986). CERES-EGC comprises sub-models for the major processes governing the cycles of water, carbon and nitrogen in soil-crop systems. A physical sub-model simulates the transfer of heat, water and nitrate down the soil profile, as well as soil evaporation, plant water uptake and transpiration in relation to climatic demand. Water infiltrates down the soil profile following a tipping-bucket approach, and may be redistributed upwards after evapo-transpiration has dried some soil layers. In both of these equations, the generalised Darcy's law has subsequently been introduced in order to better simulate water dynamics in fine-textured soils (Gabrielle et al., 1995).

A biological sub-model simulates the growth and phenology of the crops. Crop net photosynthesis is a linear function of intercepted radiation according to the Monteith approach, with interception depending on leaf are index based on Beer's law of diffusion in turbid media. Photosynthates are partitioned on a daily basis to currently growing organs (roots, leaves, stems, fruits) according to crop development stage. The latter is driven by the accumulation of growing degree days, as well as cold temperature and day-length for crops sensitive to vernalisation and photoperiod. Lastly, crop N uptake is computed through a supply/demand scheme, with soil supply depending on soil nitrate and ammonium concentrations and root length density.

A micro-biological sub-model simulates the turnover of organic matter in the plough layer. Decomposition, mineralisation and N-immobilisation are modelled with three pools of organic matter (OM): the labil OM, the microbial biomass and the humads. Kinetic rate constants define the C and N flows between the different pools.

Direct field emissions of CO_2, N_2O, NO and NH_3 into the atmosphere are simulated with different trace gas modules. Here, we focus on the nitrous oxide emissions module which is adapted from the semi-empirical model NOE (Hénault et al., 2005) for simulating the N_2O production in the soil through both the nitrification and the denitrification pathways. Denitrification component is derived from the NEMIS model (Hénault and Germon, 2000) that calculates the actual denitrification (Da) as the product of a potential rate (PDR) with three unitless factors related to soil water content (F_W), nitrate content (F_N) and temperature (F_T), as follows:

$$Da = PDR.F_N.F_W.F_T \tag{5}$$

In a similar fashion, nitrification is modelled as the product of a maximum nitrification rate (MNR) with three unitless factors related to water-filled pore space (N_W), ammonium concentration (N_N) and temperature (N_T) and expressed as follows:

$$Ni = MNR.N_N.N_W.N_T \tag{6}$$

Nitrous oxide emissions resulting from the two processes are soil-specific proportions of total denitrification and nitrification pathways and are calculated according to:

$$N_2O = r.Da + c.Ni \tag{7}$$

where r is the fraction of denitrified N and c is the fraction of nitrified N that both evolve as N_2O. The N_2O sub-model of CERES-EGC involves a total set of 15 parameters of which four of them are site-specific and must be measured on site, while the other 11 are considered global, i.e. fixed over time and space. The local parameters are the potential denitrification rate (PDR), the maximum nitrification rate (MNR) and the fractions of nitrified (c) and denitrified (r) N that evolve as N_2O. They were all measured with the same protocol as summarised by Hénault et al. (2005). For each test site, the PDR were measured by acetylene blocking of undisturbed soil cores taken from the top 20 cm of soil, satured with water and incubated with an ample supply of nitrate (Hénault and Germon, 2000). The fraction of denitrified nitrate that evolves as N_2O was determined as the difference between the N_2O production rates of soil cores incubated with and without acetylene and the fraction of nitrification evolved as N_2O was measured on sieved soil samples incubated with increasing soil moisture content and non-limiting NH_3 supply. Lastly, the maximum nitrification rate (MNR) and the fraction (c) were measured by using laboratory soil incubations and following the protocol of Garrido et al. (2002). The 11 global parameters are the constants of the N_2O module equations which are fixed from site to another and fixed in time. In the original model, the calibration of the global parameters of the denitrification process was carried out by Hénault and Germon (2000) whereas the global parameters of the nitrification pathway were calibrated by Garrido et al. (2002) and Laville et al. (2005). The different response functions computed with the default parameter values and equations presented below are shown in Figure 3 (as the dotted lines). The response functions are unitless and are calculated as:

$$F_N = \frac{\left[NO_3^-\right]}{Km_{denit} + \left[NO_3^-\right]} \tag{8}$$

where F_N is the denitrification response factor to soil nitrate content $\left[NO_3^-\right]$, in mg N kg^{-1} soil and Km_{denit} id the nitrate N content (mg N kg^{-1} soil) where $F_N = 0.5$.

$$F_W = 0,\ WFPS < Tr_{WFPS}$$

$$F_W = \left[\frac{WFPS - Tr_{WFPS}}{1 - Tr_{WFPS}}\right]^{POW},\ WFPS \geq Tr_{WFPS} \tag{9}$$

where F_W is the denitrification response factor to soil WFPS, Tr_{WFPS} is a threshold value below which no denitrification occurs and POW is the exponent of the power function.

$$F_T = exp\left[\frac{(T - TTr_{denit})\, ln\left(Q10_{denit,1}\right) - 9ln\left(Q10_{denit,2}\right)}{10}\right],\ T < TTr_{denit}$$

$$F_T = exp\left[\frac{(T-20)ln(Q10_{denit,2})}{10}\right],\ T \geq TTr_{denit} \tag{10}$$

where F_T is the denitrification response function to soil temperature (T). F_T is derived from two biological reactions that occur either below a threshold temperature TTr_{denit} or above TTr_{denit}. Both reactions present a different Q10 ($Q10_{denit,1}$ and $Q10_{denit,2}$) which are both an increase factor for a 10 °C increase in T.

$$N_N = \frac{\left[NH_4^+\right]}{Km_{nit} * H_p + \left[NH_4^+\right]} \tag{11}$$

where N_N is the nitrification response factor to soil ammonium content $\left[NH_4^+\right]$. The half-saturation constant Km_{nit} is calculated at each soil water content (Hp in w/w).

$$N_W = \frac{WFPS - MIN_{WFPS}}{OPT_{WFPS} - MIN_{WFPS}},\ MIN_{WFPS} < WFPS \leq OPT_{WFPS}$$

$$N_W = \frac{MAX_{WFPS} - WFPS}{MAX_{WFPS} - OPT_{WFPS}},\ OPT_{WFPS} \leq WFPS < MAX_{WFPS} \tag{12}$$

$$else\ N_W = 0$$

where N_W is the nitrification response function to soil water content. Nitrification is assumed to increase linearly from a minimum WFPS (MIN_{WFPS}) up to an optimal value (OPT_{WFPS}) and

then to linearly decrease down to a maximum WFPS (MAX_{WFPS}) (Rolland et al., 2008).

$$N_T = exp\left[\frac{(T-20)ln(Q10_{nit})}{10}\right] \tag{13}$$

where N_T is the response factor to soil temperature (T) and $Q10_{nit}$ is the Q10 factor for this reaction.

Once we had gathered prior information on likely ranges of variation for each parameter, based on literature review, we assigned them uniform and non-correlated probability distributions between the minimum and maximum bounds. Table 1 describes the parameters on which we applied our procedure described below.

2.4 The database of N_2O measurements

The nitrous oxide measurements were carried out on seven experiment sites which are located in Northern France (Table 2). The experiments were conducted on major arable crop types and soils types representative of the area. For some sites, different treatments were conducted with various N-fertiliser amounts supplied to the crop. Nitrous oxide emissions were monitored by the static chamber method with eight replicates for all the sites, except for the site of Grignon where measurements were monitored with three automatic chambers during 31 successive days from 13 May 2005 to 12 June 2005. Uncertainty about measurements was the standard deviation of fluxes measured with the different chambers on field. Input data required to run the model were also collected on each site. Daily weather data required by CERES-EGC (rainfall, air temperature, solar radiation) were collected from local weather stations and detailed information concerning soil parameters and crop management were compiled into input files for CERES-EGC. Uncertainty on these input data is not considered in our approach.

2.5 Our Procedure

The procedure described below was applied for two main objectives: (i) calibrate the parameters site-by-site and (ii) obtain better global estimates for the parameters initially deemed universal. The first objective was obtained by applying our Bayesian procedure for each data sample which is called the *sample-by-sample procedure*, i.e. the calibration was successively implemented for each data sample, the likelihood between the measurements and the predictions was computed using only one site-specific data set of N_2O measurements and using the associated specific model input data. The universal values of the global parameters were calibrated by running our procedure with the 11 data sets simultaneously, which is called the *multisample procedure*, i.e. by calculating the posterior distribution as:

$$p(\theta|Y_1, ..., Y_{11}) \propto p(Y_1, ..., Y_{11}|\theta)\, p(\theta) \tag{14}$$

where Y_i is the i^{th} data sample in our data base. Thus, data log-likelihood becomes equal to the sum of the log-likelihoods calculated between observations and predictions for all the data sets. The procedure of the Bayesian method we set up was built using the Metropolis-Hastings algorithm and based on running three parallel Markov chains Monte Carlo of N iterations which were started with three different parameter points (θ_0). The three starting points were defined in the prior parameter space as the default parameter values and the two lower and upper parameter bounds (θ_{min} and θ_{max}). Before running the chains, we pre-defined the chain length N, the variances matrix Σ for the proposal distribution δ and the starting points θ_0. After that, we checked to obtain a satisfactory acceptance rate (20-30%) of candidate vectors and that the chains had converged. If these two conditions were not fulfilled, we restarted new chains with different tuning values: N and Σ. Once we had run the three parallel chains, we applied the convergence diagnostic proposed by Gelman and Rubin which is based on the comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance (Gelman and Rubin,

1992). The convergence is diagnosticed when variance between the multiple chains is no larger than the variance within each individual chain. Running this test facilitates both the testing of the convergence and the identification of the convergence point. We considered that the chains had converged when the ratio of within-chains variances over the between chains variances (the Gelman and Rubin's shrink factor) approached 1. On the contrary, values substantially above 1 indicated lack of convergence. The Markov chains resulting from the random walk of the Metropolis-Hastings algorithm are highly auto-correlated in time because each iteration depends on the previous one and this is a problem to approximate the posterior pdf. In fact, the time serie with the maximum of posterior information is such as each iteration should be an independent sample from the posterior distribution (Plummer et al., 2006). For obtaining such chains less auto-correlated, the chains must be made thinner to extract independent iterations. The process of thinning was made in two steps: the auto-correlation was first computed for increasing lags and then the posterior chain was extracted by keeping the iterations defined by the thinning interval. We defined this one as the number of iterations between consecutive samples in a chain for which the auto-correlation was less than 60%. We also extracted the final subset of iterations by removing the burn in period, i.e. 10% of the iterations at the beginning of the chain. This sample can be used to summarise the posterior pdf by calculating the mean vector, the variance matrix and the 90% credible interval for each parameter. In addition, the posterior pdf facilitates the calculation of cross-correlations between parameters and the distribution of model predictions directly computed with the posterior parameter sample.

The generation of the Markov chains and the chain analysis were carried out by using the R software devices (R Development Core Team, 2008). The CERES-EGC model coded in Fortran was compiled as a R-library and called from the R software with the different parameter values generated in the loop of the Metropolis-Hastings algorithm which was coded as a R function. The chains analysis and diagnostics were carried out with the *coda* R package (Plummer et al.,

2006).

2.6 Evaluation of model predictions

The performance of the calibration procedures was assessed by calculating the root mean square error (RMSE) of predictions computed with the simulations $f(\theta)$ for which the vector θ either was drawn from the prior parameter pdfs, or was the posterior MCMC chains from sample-by-sample or multisample calibration procedure. Otherwise, θ was either a single value equal to the default parameter vector ($\theta_{default}$), or in both following cases, three parameter vectors, equal either to the mean of calibrated parameter ($\overline{\theta}$) for the 3 parallel chains or to the maximum a posteriori estimate of θ (θ_{MAP}). θ_{MAP} is the single best value of the parameter vector in each MCMC chain, at which the posterior probability distribution is maximal (Van Oijen et al., 2005). RMSE was defined as follows:

$$RMSE = \left(E\left[(O_i - S_i)^2\right]\right)^{1/2} \tag{15}$$

where O_i and S_i are the time series of the observed and the simulated data and E denotes the expectancy (Smith et al., 1996). Simulations S_i were achieved with a single run of the model in the case of the single value of $\theta_{default}$. Simulations were achieved with the 3 values of $\overline{\theta}$ and θ_{MAP} corresponding to the results from the 3 three parallel MCMC chains, then, in these cases, S_i was the expectancy of the 3 simulated time-series. In the case of prior parameter pdfs, S_i was defined as the expectancy of the simulations computed with 100 parameter vectors generated as random deviates in the prior uniform distributions. In the case of MCMC chains of parameters, S_i was the expectancy of the simulations with the thinned chains of parameters from both sample-by-sample and multisample calibration procedures.

3 Results

3.1 Posterior parameter distributions

Figure 1 portrays, in the form of boxplots, the posterior parameter distributions obtained with the application of our procedure performed sample-by-sample and with the multisample procedure. In both cases, three Markov chains were run to assure the convergence of the algorithm, then, the chains were thinned with the method described in Section 2.5 and, finally, the boxplots are the distributions of the combination of the three thinned Markov chains for each calibration. The boxplots depict the parameter distribution through the five following numbers: the median, the lower quartile (Q1), the upper quartile (Q3) and the largest observations (without outliers). This representation allows us to display differences between parameter calibration in relation to field-site and the shape of the boxplot portrays the dispersion and symmetry of the distribution. The y-axis is limited with the minimum and maximum bounds of the prior parameter distribution. Our Bayesian procedure generally generated uni-modal distributions, clearly suggesting that the MCMC chains converged towards a unique convergence point. This result was clearly corroborated by the convergence test that we systematically applied to the three parallel Markov chains. Figure 2 displays the 50 and 97.5% quantiles of the sampling distribution for the Gelman-Rubin shrink factor for the 11 parameters specifically calibrated with the data set of La Saussaye and indicates that the shrink factor approached 1 for all the parameters after 30 000 iterations which supports the convergence of the three chains for all the parameters.

Figure 1 shows that the posterior distributions become narrower compared to the uniform prior distributions which is undoubtedly due to the efficiency of our procedure. From prior uniform distribution function, the posterior parameter distribution was converging into normal or log normal distribution functions as already observed by Svensson et al. (2008) for the Bayesian calibration of a process-based forest model with a similar procedure. For example, the distribu-

tions of parameter $\theta(1)$ which is the WFPS threshold for denitrification activation appear tight on a specific value for each data sample, suggesting that the calibration have clearly reduced the uncertainty about the value of this parameter. On contrary, the parameters $\theta(8)$ and $\theta(9)$ which respectively correspond to the minimum and maximum water-filled pore space for which nitrification is activated in the topsoil layer are relatively spread in the prior range of variation and the median is rather centred in the distribution.

Figure 1 also shows that some parameter distributions are flattened on the limits of the prior bounds. We therefore should reconsider the prior ranges for these parameters, particularly the distribution of the parameters $\theta(10)$, the half-saturation constant of the ammonium function, and $\theta(11)$, the Q10 factor for nitrification, which are flattened on the minimum limit of the prior range for the data samples of Champnoël AN, La Saussaye and Grignon.

The last boxplot in the 11 graphs of Figure 1 displays the distribution obtained with the multisample procedure. The shape of this boxplot and its median value appear to be more constrained by certain data samples than others which can be explained by the fact that both data samples with a larger number of observations and individual observations with higher precision have substantially more weight in the log-likelihood function. For example, the boxplots of the multisample procedure show a similarity with the boxplots of the site La Saussaye, particularly for the parameters $\theta(1)$, $\theta(3)$ and $\theta(6)$.

Data samples acquired in the same sites, i.e. sites with both identical climate and soil type but with differentiated crop management, may generate similar shapes of distribution. The three treatments Rafidin N0, Rafidin N1, Rafidin N2, the two treatments Champnoël CT and Champnoël AN, and the two treatments Le Rheu CT and Le Rheu AN may present similar distributions for some parameters which could support the conclusion that the parameters are site-specific. The uncertainty on the parameters computed with the multisample procedure is the synthesis of the various observations and the best compromise in relation to our current obser-

vations. The posterior pdfs computed with this procedure could be useful to simulate N_2O emissions in new locations where no measurement is available for parameter calibration, whereas it could be more interesting to use site-specific calibrated parameters in the case of simulations with similar characteristics of soil to the soil types of our database.

Figure 3 demonstrates that the calibration procedure involves response functions with different shapes that result from the functions plotted with the various calibrated parameter sets. Difference between the individual functions may be quite considerable. The response functions N_N (Fig. 3.a) are very different between each other and reflect the value of $\theta(10)$. The function calibrated with the multisample procedure (dashed line in Fig. 3.a) is below the function computed with default parameter values, which means that the calibrated function reduces more the potential of nitrification rate than the function by default. The water functions N_W (Fig. 3.b) shows that the minimum WFPS for activation of the nitrification ($\theta(8)$) is centred on the default value, that the optimum WFPS for nitrification ($\theta(7)$) is lower for the calibration with data sets Le Rheu AN and La Saussaye, and that the value from multisample procedure is similar to the default parameter value. The calibrated maximum WFPS for nitrification ($\theta(9)$) are always higher than the default value and are centred on 90% WFPS which means that the nitrification could occur for higher WFPS than initially. The shapes of the response function N_T (Fig. 3.c) are similar to the initial shape for the sites La Saussaye and Grignon, and for the other sites, the shapes are equivalent between each other at least down to 25 °C. The response functions F_N (Fig. 3.d) are clearly below the function by default what suggests that the default value of $\theta(2)$ was apparently too low. The function F_W which reflects the effect of WFPS on denitrification presents a large variety of shapes depending on the parameter $\theta(1)$, the threshold of WFPS from which denitrification starts to occur, and $\theta(6)$, the exponent of the function. Hénault and Germon (2000) and Heinen (2006) showed that denitrification process was highly sensitive to $\theta(1)$ and that this parameter was really dependant on soil type. The value of $\theta(1)$ calibrated with the

multisample procedure is higher than the default value which means that the denitrification could start to occur with higher WFPS than with the default value. As proposed by Parton et al. (1996) to differentiate F_W with soil texture and as recommended by Hénault and Germon (2000), it will be preferable to recalibrate the parameters of F_W for each new soil with new experimental data. The shape of the functions F_T is quite similar, down to 25 °C, for each parameter set which lead us to believe that the function calibrated with the multisample procedure could be universally used in the future.

Bayesian calibration facilitates quantification of correlations between the calibrated parameters. All of them were cross-correlated to others and 6 of them showed correlation higher than 0.4 (Table 1), which we interpreted to mean that the different response functions of the N_2O module of CERES-EGC are coupled between each other and that the different parameters might be imagined as clusters of parameters such as suggested by Svensson et al. (2008). The parameters $\theta(1)$ and $\theta(2)$ are positively correlated and are both negatively correlated to $\theta(6)$ which suggests a coupling between the nitrate (F_N) and water (F_W) functions.

3.2 Error of prediction of the calibrated model

Table 3 summarises the RMSEs of prediction obtained with various parameters sets. Simulations were carried out as explained in Section 2.6 for each of the 11 data samples. RMSE computation facilitates comparison of model performance between the different sites and the different parameter values. RMSE based on posterior pdf is always lower than RMSE based on prior pdf excepted for the data sample of the Arrou site. RMSE was improved by 73% on average for all the data samples and the higher efficiency was observed for the site of La Saussaye with a RMSE declining by 98% when comparing simulations based on prior and posterior pdfs. Calibrations with the data samples: Grignon, Le Rheu CT, Le Rheu AN, Champnoël CT, Champnoël AN, Rafidin N0, Rafidin N1 and Rafidin N2 also were significantly efficient inducing the reduction of

RMSE between 79 and 96% by comparing predictions based on prior and posterior pdfs. Mean of RMSE for all the data samples dropped from 39 down to 6 g N_2O-N ha^{-1} d^{-1}. In the same way, RMSE declined by 41% on average when comparing simulations based on default parameter set and posterior pdf, mean of RMSE decreasing from 13 down to 6 g N_2O-N ha^{-1} d^{-1}. RMSE of prediction with posterior pdf was only higher than RMSE of prediction with default parameter values for the data sample of Champnoël CT for which RMSE slightly rose from 0.9 up to 1.4 g N_2O-N ha^{-1} d^{-1}.

RMSE comparison between simulations based on mean vector of parameters chains, prior pdf and default parameter values gave similar results to the performance of the posterior-pdf-based simulations. RMSE with predictions based on $\overline{\theta}$ was equal to 6 g N_2O-N ha^{-1} d^{-1} on average for all the sites, i.e. the same as posterior-pdf-based prediction. Hence, this result proved that the mean parameters could reasonably be used for future simulations of the sites of our database or for sites with similar soil types. The RMSE value based on simulations with θ_{MAP} was logically the lowest value of RMSE for the predictions based on the various parameter sets.

RMSE based on parameter sets of the multisample procedure decreased on average by 33% compared to RMSE based on simulations with prior pdfs and 14% compared to simulations with default parameter values which would lead us to believe that the parameter set summarised in Table 1 could be a good compromise when the model will be apply for a new site.

In addition, Table 3 shows that the calibration did not really improved the prediction for the two data samples of Villamblain and Arrou which may be explain for the first site by an uncertainty of prediction which was perhaps not due to parameter uncertainty and for the second site by an inaptitude of the model to simulate such hydromorphic soil. In fact, this result is in agreement with Hénault et al. (2005), Gabrielle et al. (2006a) and Heinen (2006) who also demonstrated that the N_2O module and particularly the sub-module of denitrification (Eqs. 5, 8, 9, 10) was not able to reproduce fitting dynamic of denitrification rate or N_2O emissions for soil with high

degree of water saturation.

3.3 Model prediction uncertainty

Simulations of N_2O emissions based on MCMC chains of parameters were generated as statistical distributions around a mean value which is traced in Figure 4 and displays the averaged temporal dynamic of daily N_2O emissions for every sites (Fig. 4.a to 4.k). This is a large benefit of the Bayesian approach because it facilitates analysis of the model output uncertainty due to the uncertainty about parameters values.

Model predictions were not still in agreement with all measurements which was due to uncertainty in both the measurements and the model. Measurement points with high uncertainty have less weight in the log likelihood function and then in the posterior probability, therefore, parameter sets that may give good agreement between measurements and predictions might not induced a such high log-likelihood. For example, the two higher N_2O flux measurements of the site of Villamblain (Fig. 4.a) have a large uncertainty which did not induce a strong constraint for the calibration, whereas various lower N_2O fluxes with lower uncertainty were more constraining in the calibration. The same remark is supported for the site of Arrou (Fig. 4.b). For the data sample of Champnoël AN (Fig. 4.e), a high peak flux of N_2O was observed in autumn that the model could not predict whereas fluxes lower than 10 g N_2O-N ha^{-1} d^{-1} were all well predicted. For the site of Grignon (Fig. 4.h), the observation points were concentrated on successive 31 daily measurements from 13 May 2005 to 12 June 2005 and reproduced a high peak flux of N_2O. Model with default parameter set ($\theta_{default}$) simulated three peaks of N_2O that were not observed in the field (results not shown, see Lehuger et al. (2007)). The first peak flux of N_2O occurred four days after the application of N fertiliser, in response to rainfall and high soil N content in the 0-30 cm topsoil layer. Two additional peak fluxes were simulated by the model during the measurement period as a consequence of two rainfall events, high nitrate content in soil and WFPS

predicted by the model greater than the WFPS threshold for denitrification ($\theta(1)$) which had been fixed to 62%. The Bayesian procedure applied on the Grignon site eliminates the simulation of the two additional peak fluxes simulated with the default parameter set what may be explained by the fact that the WFPS threshold for denitrification ($\theta(1)$) rose up to 73% which is the highest value in all the calibrations (Fig. 1.a). The calibration procedure for this specific data sample produced calibrated parameter values that eliminated all the other possible peak fluxes in the year (Fig. 4.h). For the data sample of Rafidin N0 (Fig. 4.i), observations also were concentrated on two short periods but with few observation points and inversely to Grignon, the calibration of this site highly constrained the model during this period but was less constraining out of the period of measurement.

Table 4 summarises the annual N_2O emissions for the different sites and treatments. The emissions were integrated over one year-period and the uncertainty of N_2O predictions was expressed as the 0.90 credible interval. Mean annual simulated fluxes were comprised between 88 and 3672 g N_2O-N ha^{-1} y^{-1} through the different sites and treatments and uncertainty about model prediction was quite large, notably for the sites where the predictions were high. The conversion factor is equal to the ratio of the integrated flux over the N fertiliser amount, whereas the emission factor for the application of N fertiliser is relative to "background" emissions, i.e. it is calculated as the difference between the mean cumulative N_2O-N emissions (g N_2O-N ha^{-1} y^{-1}) of fertilised and unfertilised predictions, over the amount of total N-fertiliser applied over one year. The emission factors were comprised between 0.05 and 1.12%, and the mean emission factor was 0.26%. This value is four times lower than the default value recommended by the IPCC tier 1 methodology which presents an uncertainty range of variation comprised between 0.3 and 3% (IPCC, 2006).

4 Discussion

4.1 Bayesian calibration

Our principal goal was to demonstrate the potential of a Bayesian-style calibration procedure to quantify parameter uncertainty and to reduce the uncertainties about the model. Our procedure sought to calibrate model parameters either successively sample-by-sample in order to improve model prediction for the specific sites of our database or simultaneously with all the data samples in order to find universal parameter values which could be apply for new soil conditions and spatial extrapolation of the model. This paper also aimed at simulating N_2O emissions with uncertainty quantification for different sites in Northern France which represent major soil, climate and crop management conditions. It has already been suggested that simple process-based models such as the N_2O module of CERES-EGC needs to be parameterised for each new location (Heinen, 2006). The application of our Bayesian procedure proves that is the case but our multi-sample procedure also demonstrates that it makes possible to find universal parameter values by encompassing all our current observations.

Our procedure which implies running three MCMC chains to ensure convergence of the algorithm for each data sample demonstrates that the parameter pdfs were considerably narrowed in comparison with the prior pdfs and proves that the uncertainty about parameters strongly decreased. The application of the sample-by-sample procedure also shows that the values of parameters could differ for each location (Fig. 1) inducing differentiated response functions according to the site (Fig. 3). Our results are in agreement with those of Heinen (2006) that revealed it seemed impossible to formulate universal reduction functions for the denitrification sub-module (Eqs. 5, 8, 9, 10) for the major soil types sand, loam and peat because the functions differed considerably within the soil types. We support the same conclusion but furthermore, our multisample procedure allows us to find the best universal compromise for parameter values.

The RMSE of prediction calculated with calibrated model was significantly reduced in comparison with prediction based on prior parameter values: 73% reduction on average with the sample-by-sample procedure and 33% reduction on average with the multisample procedure. These results clearly suggest that our calibration procedure has dramatically reduced the model uncertainty. In addition, the parameter vector equal to the mean of the thinned MCMC chains of parameters can easily be used to apply the model in similar soil conditions.

The parameterisation does not induce a perfect match between predictions and measurements for the temporal dynamics of daily N_2O fluxes. Measured data with high uncertainty were in particular less well predicted because they presented a high spatial variability and consequently were less constraining in the calculation of the log-likelihood function. Heinen (2006) also showed with a different calibration method that the optimised denitrification sub-module did not result in perfect prediction at the point scale.

Since the work of Van Oijen et al. (2005), it has been assumed that Bayesian calibration can be applied to process-based forest models and to specific sub-model (Svensson et al., 2008; Klemedtsson et al., 2007) and that the Metropolis-Hastings algorithm is particularly well adapted for calibration of ecosystem models with high number of parameters (Makowski et al., 2002). As we have just seen, our Bayesian calibration can be applied for agro-ecosystem models and for its specific modules by using data sets of measurements from different locations. Indeed, we have shown that uncertainty about parameters was considerably reduced and model performance was improved. Furthermore, our procedure to analyse the outputs for MCMC chains clearly advances the diagnostic on the parameter calibration. Convergence test on the three parallel chains of parameters and thinning for dealing with high auto-correlation in MCMC chains have been applied for each data set of our database and we are now convinced that this procedure improves the quality of parameterisation. We have also established a database of N_2O emissions for Northern France and in the future, it will be interesting to use this one to parameterise other models

or to compare the performance of different N_2O emissions process-based module integrated in CERES-EGC. Another direction could also be to use other kind of output data to parameterise specific module, for example the use of NO emission measurements for calibration of the nitrification sub-module (Eqs. 6, 11, 12, 13) of CERES-EGC.

4.2 Improvement of the model performance after integrating N_2O measurements

Heinen (2006) showed that the denitrification module (Eqs 5, 8 ,9, 10) needs to be parameterised for each location where the model would be used. In view of our results based on the sample-by-sample procedure for the complete N_2O module of CERES-EGC, we assume the same conclusion because the calibrated response functions differs from site to another which would lead us to believe that the model would not be universally applicable to simulate N_2O emissions by using the default parameters values without parameterisation. But, on the basis of the results from the application of the multisample procedure, we have seen that it becomes possible to find universal parameter values which integrate all our current observations. With the information contained in the parameter pdfs, there is now a higher likelihood to use these parameters in comparison with the prior pdfs or the default parameter values when the model needs to be applied for a new site. The power of a model is that it is able to predict variables over time and space but prior to its application, we must ensure that parameterisation is accurate. In fact, we would now point out that when the model will be applied to a new site without any available N_2O measurements, firstly we should check if the parameters will have already been parameterised for the same soil type, if yes, we could use the parameters specifically-calibrated for this soil type, else, it will be possible to use the parameters values which will have been calibrated with the multisample procedure. Indeed, it becomes possible to imagine that the parameter values from our both procedures could be used for spatial extrapolation of the model at the regional scale. In the future, as soon as a

new data set will be available we could assimilate the new observations points with the multi-sample procedure in order to reduce uncertainty about the global parameters and to increase the universality of the model.

4.3 Prediction of N_2O fluxes from agro-ecosystems

CERES-EGC and its specific N_2O module have widely been used in many soil conditions (Hénault et al., 2005; Dambreville et al., 2008; Heinen, 2006) and the model uncertainty has only been really quantified once by Gabrielle et al. (2006a) by using 5 different uncertain parameters. Likewise, uncertainty analysis about parameters is rarely carried out for process-based ecosystem models that simulate N_2O emissions. As we have just seen our Bayesian calibration resulted in probabilistic simulation of temporal dynamic of N_2O emissions over cropping seasons including the information about the probability of parameters. The calibrated model could predict temporal dynamic of daily instantaneous N_2O fluxes rather well, except for the highest peaks with high uncertainty for which the model did not perform to well predict them. In addition, the procedure makes it possible to quantify model output uncertainty for yearly N_2O budget and emission factors (EFs). Model prediction of yearly cumulative N_2O fluxes were comprised between 88 and 3672 g N_2O-N $ha^{-1}y^{-1}$ over the different sites and EFs ranged from 0.05 to 1.12%. On the basis of these results alongside those of Gabrielle et al. (2006a), it now appears that the IPCC EF is not suitable for the sites that we have studied because it considerably overestimates the annual emissions (Table 4). Beheydt et al. (2007) used the DNDC model to calculate EF corresponding to various scenarios of simulations with high N input levels and N surplus, and they found EF predicted by DNDC equal to 6.49% a value 25 times higher than our prediction and EF derived from the measurements equal to 3.16%. In the future, we also should assimilate data sets with higher level of N_2O emissions such as those used by Beheydt et al. (2007) in order to calibrate the model with annual emissions higher than 10 kg N_2O-N ha^{-1} y^{-1}. Furthermore, our results sug-

gest that the annual N_2O emissions were not strictly proportional to the application of N fertiliser which is in agreement with the results of Barton et al. (2008). Indeed, they have showed that, in a semi-arid climate, in spite of the application of N fertiliser the annual N_2O emissions were not significantly increased in comparison with background emissions and demonstrated that the N_2O emissions from arable soils can not be directly derived from the application of N fertiliser. In light of these results, we are now of the opinion that our Bayesian procedure is highly informative about model uncertainty quantification and can be very useful for taking into account risk in model-based GWP quantification of agro-ecosystems, environmental balance assessment of cropping systems and decision-making. Nevertheless, we should advice, for an efficient calibration of N_2O emissions models, that N_2O measurements with the static chamber method should be carried out with a regular recurrence at the yearly scale, at least one observation per month, and with a higher frequency during the peak fluxes that occur after N-fertiliser inputs and events that activate mineralisation of crop residues during autumn.

In the future, it would be very interesting to compare performance of various agro-ecosystem models for their aptitude to predict N_2O emissions on the same data sets in the fashion of Frolking et al. (1998); Li et al. (2005). Furthermore, Bayesian Model Comparison (Van Oijen et al., 2005; Kass and Raftery, 1995) could be applied to examine multiple models and to quantify their relative likelihood, i.e. by determining which model is most probable in view of the data and prior information.

4.4 Limits and developments of CERES-EGC to predict N_2O emissions

Our uncertainty analysis was performed without taking into account for uncertainty about the input variables of the N_2O sub-model which are the soil temperature, the soil nitrate and ammonium concentrations and the soil water content. These variables are daily calculated by the model and are dependant of numerous process (crop N uptake, nitrate leaching, evapo-transpiration,

drainage, N gas emissions, soil organic matter turnover) and thus of a large number of parameters and variables interacting over time. The Bayesian calibration could be expanded to multiple other parameters by using measured data which might be crop biomass, soil N concentration, soil water content, soil temperature and other gas emissions. In fact, the Bayesian technique could be used in a more holistic way because the method gives the possibility of calibrating various parameters by using different kind of output data (Klemedtsson et al., 2007).

We have seen that the 11 global parameters we studied depend on soil type and hence they are variable over space, as a consequence, they seem themselves to be relied on other parameters that contain a prior inherent variability. In the future, we believe that it could be very peculiar to define these new hyperparameters which control the spatial variability of our parameters (Clark, 2005) and very interesting to deal with this spatial variability by developing a hierarchical Bayesian approach. Hence, the use of our plot-scale measurements could allow us to properly extrapolate our model at the regional scale.

Another future development of the N_2O module of CERES-EGC, indispensable for model extrapolation and already mentioned by Hénault et al. (2005), should be to render the model independent of the four local parameters which still need to be measured for new soil conditions. It will require to define the key controls of these parameters in order to model them in relation to soil type characteristics.

5 Conclusion

We have shown that Bayesian calibration was successfully applied to the CERES-EGC agro-ecosystem model in order to parameterise its N_2O emissions module. We have demonstrated that both the Bayesian calibration and our procedure of analysis and diagnostic for MCMC chains of parameters can be applied to our process-based model in order to calibrate parameters either (i) by using successively data samples or (ii) by using all the data samples simultaneously, to satisfy

our objectives which were, respectively, to improve model prediction at the field scale and to find universal values of parameters in order to spatially extrapolate the model. In addition, Bayesian calibration has given us the possibility of quantifying both parameter and model uncertainty. Furthermore, it appears reasonable to assume that when the model should be applied at a larger scale than the plot-scale, the parameter values resulted from the multisample procedure could then be used for soil types which will have never been parameterised. In fact, the posterior parameter distributions encompass all our current observations and give us the possibility of quantifying their uncertainty. To this end we recommend further research into modern Bayesian method which could help us to deal with the spatial variability of the parameters.

Acknowledgements

This work formed part of the NitroEurope Integrated Project (EU's Sixth Framework Programme for Research and Technological Development) which investigates about the nitrogen cycle and its influence on the European greenhouse gas balance. We wish to thank colleagues from INRA: Matieyiendu Lamboni and Hervé Monod for their useful advices and discussions. Special thank to Christophe Dambreville and Jean-Claude Germon for making available data from Brittany.

References

Barton, L., Kiese, R., Gatter, D., Butterbach-Bahl, K., Buck, R., Hinz, C., and Murphy, D. V. (2008). Nitrous oxide emissions from a cropped soil in a semi-arid climate. *Global Change Biology*, 14(1):177–192.

Bateman, E. J. and Baggs, E. M. (2005). Contributions of nitrification and denitrification to N_2O emissions from soils at different water-filled pore space. *Biology and Fertility of Soils*, 41(6):379–388.

Beheydt, D., Boeckx, P., Sleutel, S., Li, C. S., and Van Cleemput, O. (2007). Validation of DNDC for 22 long-term N_2O field emission measurements. *Atmospheric Environment*, 41(29):6196–6211.

Butterbach-Bahl, K., Kesik, M., Miehle, P., Papen, H., and Li, C. (2004). Quantifying the regional source strength of N-trace gases across agricultural and forest ecosystems with process based models. *Plant and Soil*, 260(1–2):311–329.

Chatskikh, D., Olesen, J., Berntsen, J., Regina, K., and Yamulki, S. (2005). Simulation of effects of soils, climate and management on N_2O emission from grasslands. *Biogeochemistry*, 76(3):395–419.

Clark, J. S. (2005). Why environmental scientists are becoming bayesians. *Ecology Letters*, 8(1):2–14.

Dambreville, C., Morvan, T., and Germon, J. C. (2008). N_2O emission in maize-crops fertilized with pig slurry, matured pig manure or ammonium nitrate in Brittany. *Agriculture Ecosystems & Environment*, 123(1):201–210.

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Ojima, D. S., Kulmala, A. E., and Phongpan, S. (2000). General model for N_2O and N_2 gas emissions from soils due to dentrification. *Global Biogeochemical Cycles*, 14(4):1045–1060.

Ding, W. X., Cai, Y., Cai, Z. C., Yagi, K., and Zheng, X. H. (2007). Nitrous oxide emissions from an intensively cultivated maize-wheat rotation soil in the North China Plain. *Science of the Total Environment*, 373(2–3):501–511.

Dobbie, K. E. and Smith, K. A. (2001). The effects of temperature, water-filled pore space and land use on N_2O emissions from an imperfectly drained gleysol. *European Journal of Soil Science*, 52(4):667–673.

Frolking, S. E., Mosier, A. R., Ojima, D. S., Li, C., Parton, W. J., Potter, C. S., Priesack, E., Stenger, R., Haberbosch, C., Dorsch, P., Flessa, H., and Smith, K. A. (1998). Comparison of N_2O emissions from soils at three temperate agricultural sites: simulations of year-round measurements by four models. *Nutrient Cycling in Agroecosystems*, 52(2):77–105.

Gabrielle, B. (2006). Sensitivity and uncertainty analysis of a static denitrification model. In Wallach, D., Makowski, D., and Jones, J. W., editors, *Working with dynamic crop models: evaluating, analyzing, parameterizing and using them*, chapter 14. Elsevier.

Gabrielle, B., Laville, P., Duval, O., Nicoullaud, B., Germon, J. C., and Hénault, C. (2006a). Process-based modeling of nitrous oxide emissions from wheat-cropped soils at the subregional scale. *Global Biogeochemical Cycles*, 20(4).

Gabrielle, B., Laville, P., Hénault, C., Nicoullaud, B., and Germon, J. C. (2006b). Simulation of nitrous oxide emissions from wheat-cropped soils using CERES. *Nutrient Cycling in Agroecosystems*, 74(2):133–146.

Gabrielle, B., Menasseri, S., and Houot, S. (1995). Analysis and field-evaluation of the CERES models water-balance component. *Soil Science Society of America Journal*, 59(5):1403–1412.

Gallagher, M. and Doherty, J. (2007). Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modelling & Software*, 22(7):1000–1020.

Garrido, F., Hénault, C., Gaillard, H., Perez, S., and Germon, J. C. (2002). N_2O and NO emissions by agricultural soils with low hydraulic potentials. *Soil Biology & Biochemistry*, 34(5):559–575.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Gosse, G., Cellier, P., Denoroy, P., Gabrielle, B., Laville, P., Leviel, B., Justes, E., Nicolardot, B., Mary, B., Recous, S., Germon, J. C., Hénault, C., and Leech, P. K. (1999). Water, carbon and nitrogen cycling in a rendzina soil cropped with winter oilseed rape: the Chalons Oilseed Rape Database. *Agronomie*, 19(2):119–124.

Heinen, M. (2006). Application of a widely used denitrification model to Dutch data sets. *Geoderma*, 133(3–4):464–473.

Hénault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J. C., and Cellier, P. (2005). Predicting in situ soil N_2O emission using NOE algorithm and soil database. *Global Change Biology*, 11(1):115–127.

Hénault, C. and Germon, J. C. (2000). NEMIS, a predictive model of denitrification on the field scale. *European Journal of Soil Science*, 51(2):257–270.

Hong, B. G., Strawderman, R. L., Swaney, D. P., and Weinstein, D. A. (2005). Bayesian estimation of input parameters of a nitrogen cycle model applied to a forested reference watershed, Hubbard Brook Watershed Six. *Water Resources Research*, 41(3).

IPCC (2006). *2006 IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds).* Published: IGES, Japan.

Jambert, C., Serca, D., and Delmas, R. (1997). Quantification of N-losses as NH_3, NO, N_2O and N_2 from fertilized maize fields in Southwestern France. *Nutrient Cycling in Agroecosystems*, 48(1):91–104.

Johnsson, H., Klemedtsson, L., Nilsson, A., Bo, H., and Svensson, B. (2004). Simulation of field scale denitrification losses from soils under grass ley and barley. *Plant and Soil*, 138(2):287–302.

Jones, C. A. and Kiniry, J. R. (1986). *CERES-N Maize: a simulation model of maize growth and development*. Texas A&M University Press, College Statio, Temple, TX.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kiers, E. T., Leakey, R. R. B., Izac, A. M., Heinemann, J. A., Rosenthal, E., Nathan, D., and Jiggins, J. (2008). Agriculture at a crossroads. *Science*, 320(5874):320–321.

Klemedtsson, L., Jansson, P.-E., Gustafsson, D., Karlberg, L., Weslien, P., von Arnold, K., Ernfors, M., Langvall, O., and Lindroth, A. (2007). Bayesian calibration method used to elucidate carbon turnover in forest on drained organic soil. *Biogeochemistry*.

Larssen, T., Huseby, R. B., Cosby, B. J., Host, G., Hogasen, T., and Aldrin, M. (2006). Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach. *Environmental Science & Technology*, 40(24):7841–7847.

Laville, P., Hénault, C., Gabrielle, B., and Serca, D. (2005). Measurement and modelling of NO fluxes on maize and wheat crops during their growing seasons: effect of crop management. *Nutrient Cycling in Agroecosystems*, 72(2):159–171.

Lehuger, S., Gabrielle, B., Larmanou, E., Laville, P., Cellier, P., and Loubet, B. (2007). Predicting the global warming potential of agro-ecosystems. *Biogeosciences Discussions*, 4(2):1059–1092.

Li, C. S. (2000). Modeling trace gas emissions from agricultural ecosystems. *Nutrient Cycling in Agroecosystems*, 58(1):259–276.

Li, Y., Chen, D. L., Zhang, Y. M., Edis, R., and Ding, H. (2005). Comparison of three model-

ing approaches for simulating denitrification and nitrous oxide emissions from loam-textured arable soils. *Global Biogeochemical Cycles*, 19(3).

Linn, D. M. and Doran, J. W. (1984). Effect of Water-Filled Pore Space on Carbon Dioxide and Nitrous Oxide Production in Tilled and Nontilled Soils. *Soil Sci Soc Am J*, 48(6):1267–1272.

Maag, M. and Vinther, F. P. (1996). Nitrous oxide emission by nitrification and denitrification in different soil types and at different soil moisture contents and temperatures. *Applied Soil Ecology*, 4(1):5–14.

Maag, M. and Vinther, F. P. (1999). Effect of temperature and water on gaseous emissions from soils treated with animal slurry. *Soil Science Society of America Journal*, 63(4):858–865.

Makowski, D., Wallach, D., and Tremblay, M. (2002). Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. *Agronomie*, 22(2):191–203.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Parton, W. J., Holland, E. A., Del Grosso, S. J., Hartman, M. D., Martin, R. E., Mosier, A. R., Ojima, D. S., and Schimel, D. S. (2001). Generalized model for NO_x and N_2O emissions from soils. *Journal of Geophysical Research-atmospheres*, 106(15):17403–17419.

Parton, W. J., Mosier, A. R., Ojima, D. S., Valentine, D. W., Schimel, D. S., Weier, K., and Kulmala, A. E. (1996). Generalized model for N_2 and N_2O production from nitrification and denitrification. *Global Biogeochemical Cycles*, 10(3):401–412.

Pihlatie, M., Syvasalo, E., Simojoki, A., Esala, M., and Regina, K. (2004). Contribution of

nitrification and denitrification to N_2O production in peat, clay and loamy sand soils under different soil moisture conditions. *Nutrient Cycling in Agroecosystems*, 70(2):135–141.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1):7–11.

Qian, S. S., Stow, C. A., and Borsuk, M. E. (2003). On Monte Carlo methods for Bayesian inference. *Ecological Modelling*, 159(2–3):269–277.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Renault, P., Sierra, J., and Stengel, P. (1994). Oxygen-transport and anaerobiosis in aggregated soils - contribution to the study of denitrification. *Agronomie*, 14(6):395–409.

Ricciuto, D. M., Butler, M. P., Davis, K. J., Cook, B. D., Bakwin, P. S., Andrews, A., and Teclaw, R. M. (2008). Causes of interannual variability in ecosystem-atmosphere CO_2 exchange in a northern Wisconsin forest using a Bayesian model calibration. *Agricultural and Forest Meteorology*, 148(2):309–327.

Rolland, M. N., Gabrielle, B., Laville, P., Serca, D., Cortinovis, J., Larmanou, E., Lehuger, S., and Cellier, P. (2008). Modeling of nitric oxide emissions from temperate agricultural soils. *Nutrient Cycling in Agroecosystems*, 80(1):75–93.

Skopp, J., Jawson, M. D., and Doran, J. W. (1990). Steady-State Aerobic Microbial Activity as a Function of Soil Water Content. *Soil Sci Soc Am J*, 54(6):1619–1625.

Smith, J., Smith, P., and Addiscott, T. (1996). Quantitative methods to evaluate and compare soil organic matter (SOM) models. In Powlson, D., Smith, P., and Smith, J. E., editors, *Evaluation*

of soil organic matter models using existing long-term datasets, volume 38 of *1*, pages 183–202. Springer-Verlag, Heidelberg.

Smith, K. A. (1997). The potential for feedback effects induced by global warming on emissions of nitrous oxide by soils. *Global Change Biology*, 3(4):327–338.

Smith, K. A., Thomson, P. E., Clayton, H., McTaggart, I. P., and Conen, F. (1998). Effects of temperature, water content and nitrogen fertilisation on emissions of nitrous oxide by soils. *Atmospheric Environment*, 32(19):3301–3309.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., OMara, F., Rice, C., Scholes, B., and Sirotenko, O. (2007). *Agriculture. In Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, [B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer (eds)]*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stanford, G., Dzienia, S., and Vander Pol, R. A. (1975). Effect of Temperature on Denitrification Rate in Soils. *Soil Sci Soc Am J*, 39(5):867–870.

Sutton, M. A., Nemitz, E., Erisman, J. W., Beier, C., Bahl, K. B., Cellier, P., de Vries, W., Cotrufo, F., Skiba, U., Di Marco, C., Jones, S., Laville, P., Soussana, J. F., Loubet, B., Twigg, M., Famulari, D., Whitehead, J., Gallagher, M. W., Neftel, A., Flechard, C. R., Herrmann, B., Calanca, P. L., Schjoerring, J. K., Daemmgen, U., Horvath, L., Tang, Y. S., Emmett, B. A., Tietema, A., Penuelas, J., Kesik, M., Brueggemann, N., Pilegaard, K., Vesala, T., Campbell, C. L., Olesen, J. E., Dragosits, U., Theobald, M. R., Levy, P., Mobbs, D. C., Milne, R., Viovy, N., Vuichard, N., Smith, J. U., Smith, P., Bergamaschi, P., Fowler, D., and Reis, S. (2007).

Challenges in quantifying biosphere-atmosphere exchange of nitrogen species. *Environmental Pollution*, 150(1):125–139.

Svensson, M., Jansson, P. E., Gustafsson, D., Kleja, D. B., Langvall, O., and Lindroth, A. (2008). Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand. *Ecological Modelling*, 213(3–4):331–344.

Van Oijen, M., Rougier, J., and Smith, R. (2005). Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology*, 25(7):915–927.

List of Tables

Parameter vector $\theta = [\theta(1)...\theta(11)]$					Prior probability distribution			Posterior probability distribution		
$\theta(i)$	Symbol	Description	Unit	Default value	θ_{min}(i)	θ_{max}(i)	References	Mean	SD	Correlated $\{\theta(i)\}$
$\theta(1)$	Tr_{WFPS}	WFPS threshold for denitrification	%	0.62	0.40	0.80	Gabrielle (2006); Hénault et al. (2005) Hénault and Germon (2000); Johnsson et al. (2004)	0.689	0.007	{2,6}
$\theta(2)$	Km_{denit}	Half-saturation constant (denit)	mg N kg^{-1} soil	22.00	5.00	120.00	Gabrielle (2006); Ding et al. (2007) Parton et al. (2001); Del Grosso et al. (2000) Parton et al. (1996); Bateman and Baggs (2005) Johnsson et al. (2004)	66.94	22.47	{1,6}
$\theta(3)$	TTr_{denit}	Temperature threshold	°C	11.00	10.00	15.00	Gabrielle (2006); Johnsson et al. (2004) Renault et al. (1994)	10.27	0.17	
$\theta(4)$	$Q10_{denit,1}$	Q10 factor for low temperature	Unitless	89.00	60.00	120.00	Stanford et al. (1975); Maag and Vinther (1999)	89.46	18.28	{5}
$\theta(5)$	$Q10_{denit,2}$	Q10 factor for high temperature	Unitless	2.10	1.00	4.80	Gabrielle (2006); Stanford et al. (1975)	2.62	1.17	{4,10}
$\theta(6)$	POW_{denit}	Exponent of power function	Unitless	1.74	0.00	2.00	Stanford et al. (1975); Smith et al. (1998) Johnsson et al. (2004); Maag and Vinther (1999) Maag and Vinther (1996); Skopp et al. (1990)	1.53	0.23	{1, 2}
$\theta(7)$	OPT_{WFPS}	Optimum WFPS for nitrification	%	0.60	0.35	0.75	Jambert et al. (1997); Laville et al. (2005)	0.59	0.12	
$\theta(8)$	MIN_{WFPS}	Minimum WFPS for nitrification	%	0.10	0.05	0.15	Linn and Doran (1984); Jambert et al. (1997) Skopp et al. (1990); Ding et al. (2007) Parton et al. (2001); Bateman and Baggs (2005)	0.095	0.02	
$\theta(9)$	MAX_{WFPS}	Maximum WFPS for nitrification	%	0.80	0.80	1.00	Linn and Doran (1984); Parton et al. (2001) Bateman and Baggs (2005)	0.88	0.05	
$\theta(10)$	Km_{nit}	Half-saturation constant (nit)	mg N kg^{-1} soil	10.00	1.00	50.00	Linn and Doran (1984); Jambert et al. (1997) Pihlatie et al. (2004)	25.69	14.17	{5}
$\theta(11)$	$Q10_{nit}$	Q10 factor for nitrification	Unitless	2.10	1.90	13.00	Maag and Vinther (1996); Laville et al. (2005) Smith (1997); Dobbie and Smith (2001)	7.36	3.04	

Table 1: Description of the 11 parameters of the N_2O emissions module. The prior probability distribution is defined as multivariate uniform between bounds θ_{min} and θ_{max} which were extracted from literature review. The posterior parameter distributions are based on the multisample procedure and are characterised with the mean value of parameter chains, the standard deviation (SD) and the parameters with which each parameter is correlated (underline if negative) at greater absolute value than 0.4

Site	Treatment	Year	Soil texture	Crop type	N fertiliser (kg N ha^{-1})	Number of observations	Source
Rafidin	N0	1994-1995	Rendzina	Rapeseed	0	7	Gosse et al. (1999)
	N1	1994-1995	Rendzina	Rapeseed	155	8	Gosse et al. (1999)
	N2	1994-1995	Rendzina	Rapeseed	262	9	Gosse et al. (1999)
Villamblain		1998-1999	Loamy Clay	Winter Wheat	230	15	Hénault et al. (2005)
Arrou		1998-1999	Loamy Clay	Winter Wheat	180	18	Hénault et al. (2005)
La Saussaye		1998-1999	Clay Loams	Winter Wheat	200	14	Hénault et al. (2005)
Champnoël	CT	2002-2003	Silt Loam	Maize	0	15	Dambreville et al. (2008)
	AN	2002-2003	Silt Loam	Maize	110	23	Dambreville et al. (2008)
Le Rheu	CT	2004-2005	Silt Loam	Maize	18	24	Dambreville et al. (2008)
	AN	2004-2005	Silt Loam	Maize	180	22	Dambreville et al. (2008)
Grignon		2005	Silt Loam	Maize	140	31	Lehuger et al. (2007)

Table 2: Database of N_2O emissions from different arable sites in France. For the different data sets, following characteristics are described: year of measurement, soil texture, crop type, application of N fertiliser , number of observation points and references

Site	Treatment	RMSE (in g N_2O-N ha^{-1} d^{-1}) computed with predictions based on:					
		Prior pdf	Default parameter values	Posterior pdf	Mean of parameter chain	Parameter set with maximum log-likelihood	Multisample procedure
Rafidin	N0	4.6	4.6	0.7	0.3	0.3	4.6
	N1	7.5	10.4	1.2	1.4	1.2	12.8
	N2	10.5	15.9	2.1	3.0	2.8	20.4
Villamblain		5.2	5.5	4.8	4.9	4.9	5.5
Arrou		25.4	29.0	27.1	25.3	23.8	29.2
La Saussaye		93.0	33.3	2.0	2.3	2.4	2.3
Champnoël	CT	21.5	0.9	1.4	0.9	0.9	0.9
	AN	65.58	15.0	13.8	14.0	13.8	14.0
Le Rheu	CT	149.5	22.7	6.1	6.0	6.0	6.0
	AN	30.4	4.2	2.0	2.2	2.2	2.4
Grignon		16.9	1.0	1.0	1.2	1.3	1.1

Table 3: Root mean square error (RMSE) of prediction (in g N_2O-N ha^{-1} d^{-1}) based on different parameter sets: the prior pdf, the default parameter values, the posterior pdf, the mean of MCMC chains, the parameter set with maximum log-likelihood and the mean of parameter chains from the multisample procedure

Site	Treatment	Year	N_2O Fluxes (g N ha^{-1} y^{-1})	0.05 quantile (g N ha^{-1} y^{-1})	0.95 quantile (g N ha^{-1} y^{-1})	IPCC (g N ha^{-1} y^{-1})	Conversion factor (%)	Emission factor (%)
Rafidin	N0	1994-1995	689	578	741	0	-	-
	N1	1994-1995	584	473	824	1550	0.4 (0.3-0.5)	0.07 (0.00-0.22)
	N2	1994-1995	819	629	1183	2620	0.3 (0.2-0.5)	0.10 (0.03-0.24)
Villamblain		1998-1999	1465	454	2989	2300	0.6 (0.2-1.3)	0.36 (0.00-1.02)
Arrou		1998-1999	3672	1676	5874	1800	2.0 (0.9-3.3)	0.26 (0.00-1.49)
La Saussaye		1998-1999	3215	572	6035	2000	1.6 (0.3-3.0)	1.12 (0.00-2.53)
Champnoël	CT	2002-2003	218	49	746	0	-	-
	AN	2002-2003	336	106	855	1100	0.3 (0.1-0.8)	0.06 (0.00-0.53)
Le Rheu	CT	2004-2005	88	66	115	180	0.5 (0.4-0.6)	-
	AN	2004-2005	183	146	220	1800	0.10 (0.08-0.12)	0.05 (0.03-0.08)
Grignon		2005-2006	150	143	163	1400	0.11 (0.10-0.12)	0.05 (0.04-0.05)

Table 4: Cumulative annual N_2O fluxes (g N_2O-N $ha^{-1}y^{-1}$) computed as the sum of mean, 0.05 quantile and 0.95 quantile daily simulations based on parameter MCMC thinned chains. Annual estimates from IPCC methodology (corresponding to the emissions due to fertiliser application), conversion factor (%) and emission factor (%) with range of uncertainty were also reported

List of Figures

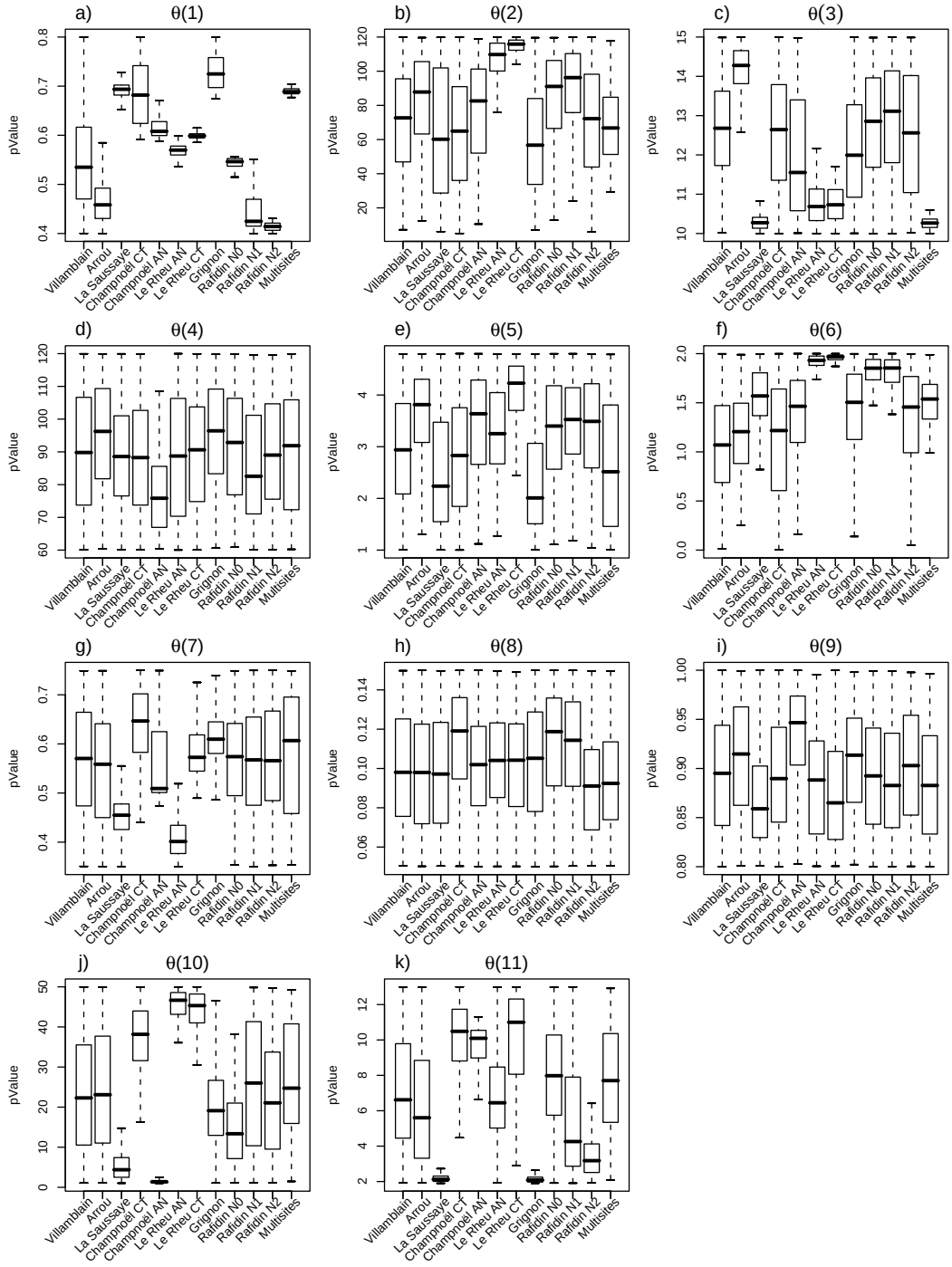

Figure 1: Posterior distribution of the 11 calibrated parameters ($\theta(1)$ to $\theta(11)$) represented as boxplots over the prior range of variation. For each graph, the boxplots are computed from calibration sample-by-sample and with the "multisample" procedure

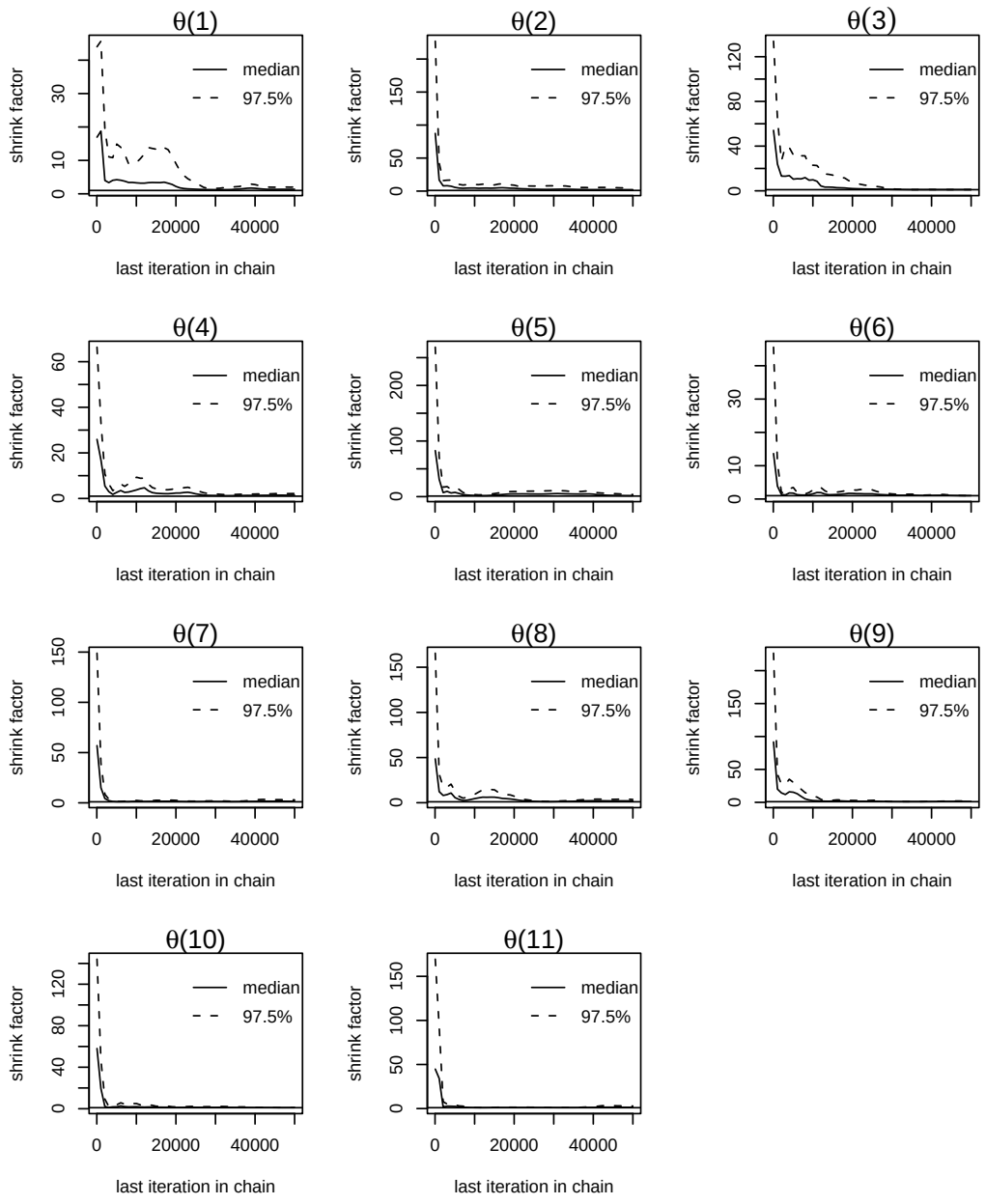

Figure 2: Evolution of the Gelman and Rubin's shrink factor for the calibration of the site La Saussaye

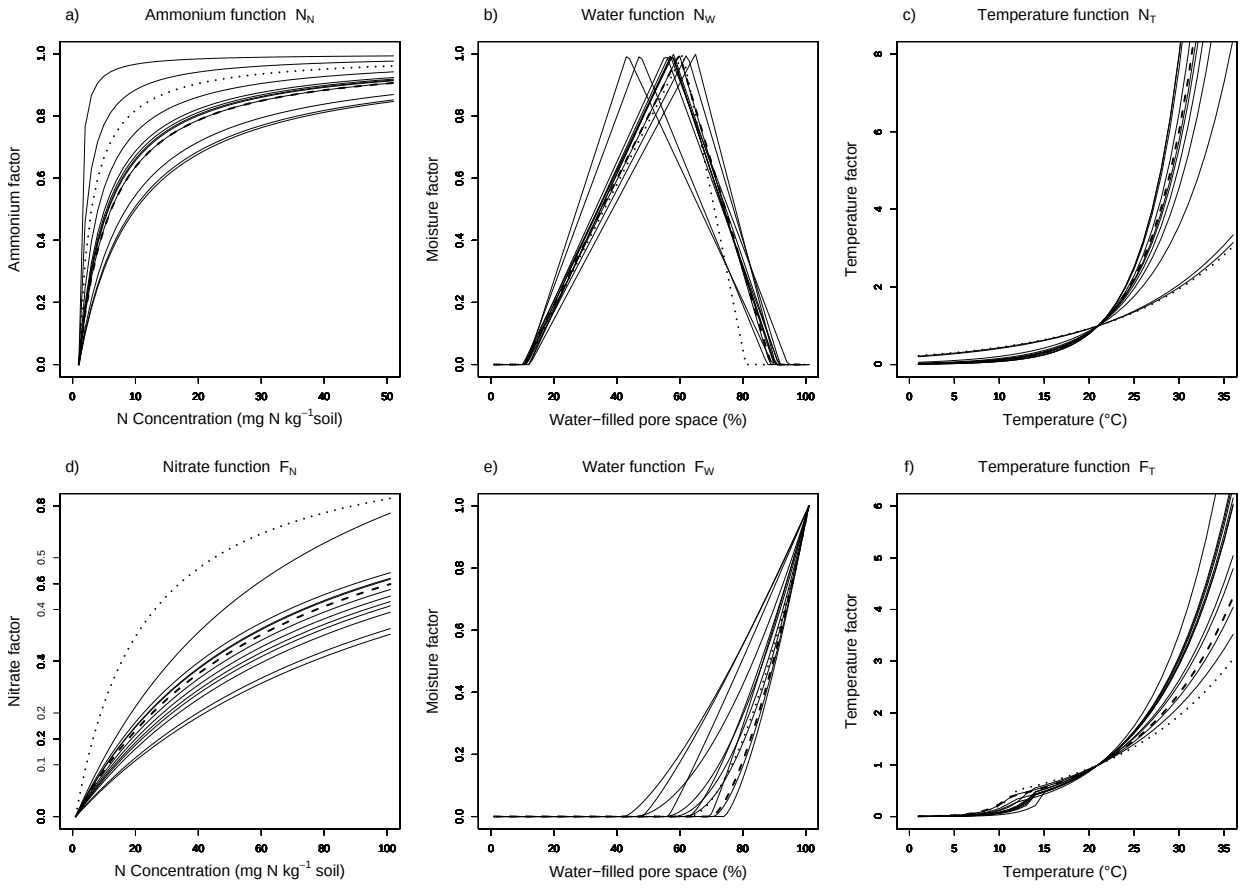

Figure 3: Response functions of the N_2O emission module traced with different parameters sets: default parameter values (dotted line), mean of parameters chains for each sample-by-sample calibration (line) and with mean of parameter chains for the multisample calibration (dashed line)

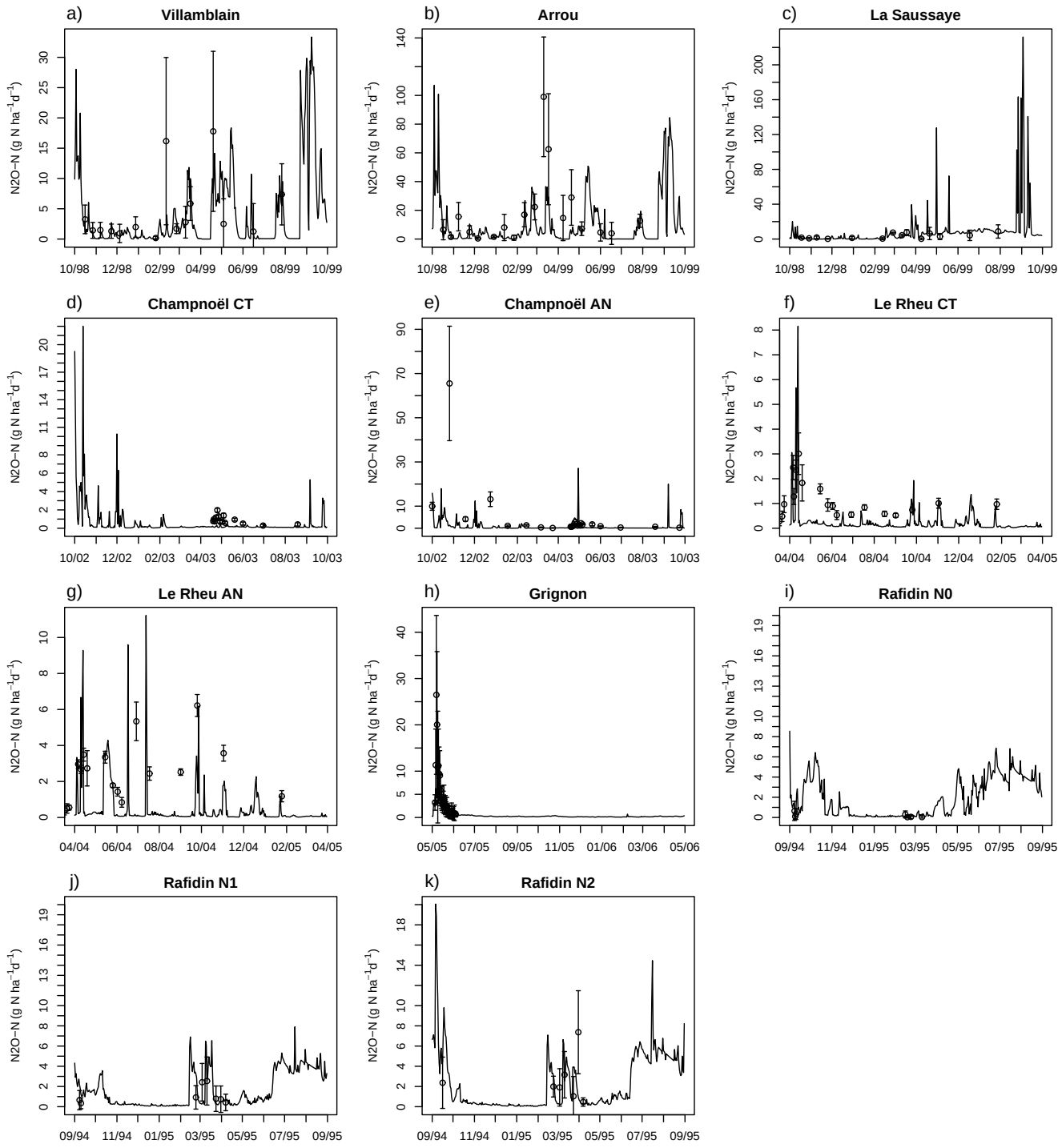

Figure 4: Simulated (lines) and observed (symbols) N_2O emissions for the different sites and treatments. The simulated line is the mean time serie of the simulations based on the MCMC chains of parameters from sample-by-sample calibrations