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 N_2O emissions in relation to environmental conditions and crop management. Biophysical models simulating the dynamics of carbon and nitrogen in agro-ecosystems have a unique potential to explore these relationships, but are fraught with high uncertainties in their parameters due to their variations over time and space. Here, we used a Bayesian approach to calibrate the parameters of the N_2O submodel of the agro-ecosystem model CERES-EGC. The submodel simulates N_2O emissions from the nitrification and denitrification processes, which are modelled as the product of a potential rate with three dimensionless factors related to soil water content, nitrogen content and temperature. These equations involve a total set of 15 parameters, four of which are site-specific and should be measured on site, while the other 11 are considered global, i.e. invariant over time and space. We first 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 developed to update the parameter distributions against a database of seven different field-sites in France. Three parallel Markov chains were run to ensure a convergence of the algorithm. This site-specific calibration significantly reduced the spread in parameter distribution, and the uncertainty in the N_2O simulations. The model's root mean square error (RMSE) was also abated by 73% across the field sites compared to the prior parameterization. The Bayesian calibration was subsequently applied simultaneously to all data sets, 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; Greenhouse gases

1 Introduction

Soils are the main source of nitrous oxide (N_2O) in the atmosphere, via the microbial processes of nitrification and denitrification. Because of its heavy reliance on synthetic N-fertilisers, agriculture has enhanced these two processes, as a result of which agro-ecosystems contribute 55-65% of the global anthropogenic emissions of N_2O. Compared to other ecosystem types or economic sectors, they are thus responsible for the major part of the atmospheric build-up of N_2O (Smith et al., 2007). Compared to other greenhouse gases (GHG) such as CO_2, N_2O fluxes are of small magnitude and highly variable in space and time, being tightly linked to the local climatic sequence and soil properties. Predicting N_2O emissions from agro-ecosystems thus requires taking into account complex processes and interactions which originate from both environmental conditions and agricultural practises (Duxbury and Bouldin, 1982; Grant and Pattey, 2003; Pattey et al., 2007). This poses a serious challenge to the estimation of the source strength of arable soils, which is currently mostly based on available statistics on fertilizer ignoring these environmental factors (IPCC, 2006; Lokupitiya and Paustian, 2006). On the other hand, process-based agro-ecosystem models may in principle capture these effects, and have thereby a unique potential to predict N_2O emissions from arable soils at the plot-scale as well as at regional and continental scales (Butterbach-Bahl et al., 2004; Li et al., 2001; Gabrielle et al., 2006a; Del Grosso et al., 2006). Examples of biophysical N_2O-models include DAYCENT (Parton et al., 2001), DNDC (Li, 2000), FASSET (Chatskikh et al., 2005) and CERES-EGC (Gabrielle et al., 2006b). However, a major limitation to the wide-spread use of these models lies in the fact that their predictions are highly dependent on parameter settings, and carry a large uncertainty due to uncertainties in parameter values, driving variables and model structure (Gabrielle et al., 2006a). Although model parameterisation and uncertainty analysis are widely developed in the literature on agro-ecosystem models, they are rarely considered simultaneously (Monod et al., 2006;

Makowski et al., 2006). Bayesian calibration makes it possible to combine the two types of analysis by providing estimates of parameters values under the form of probability density functions (pdfs), which may be also propagated to model outputs 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 (Makowski et al., 2006).

In ecological and environmental sciences, Bayesian calibration has been applied to a wide range of models (Hong et al., 2005; Larssen et al., 2006; Ricciuto et al., 2008), and this field is developing actively, mainly using Markov Chain Monte Carlo (MCMC) methods to estimate the posterior pdf for the model parameters. The Bayesian methodology described by Van Oijen et al. (2005) was applied to dynamic process-based forest models with the objective of calibrating model parameters with various types of observed data from forested experimental sites (Svensson et al., 2008; Klemedtsson et al., 2007). In these examples, Metropolis-Hastings MCMC-algorithm was used to generate samples from the posterior parameter distributions. Although there is an increasing body of literature on the application of Bayesian approaches to environmental sciences, the latter have not been applied to process-based model of soil N_2O emission models, to the best of our knowledge.

The overall purpose of this paper was thus to calibrate the parameters of the N_2O emission module of the CERES-EGC agro-ecosystem model and to quantify uncertainty of model simulations by developing a suitable Bayesian calibration method. Data sets of measured N_2O emission rates were collected from seven field-sites in Northern France, which represent major soil types, crops and management practices of the area. The Bayesian procedure was first applied separately to each experimental site, and secondly to the ensemble of the sites. This made it possible to explore the spatial variability of model parameters, and to test whether they could be considered as universal and with which uncertainty range.

2 Material and Methods

We carried out Bayesian calibration using the Metropolis-Hastings algorithm, to estimate the joint probability distribution for the parameters of the N_2O emission module of the CERES-EGC model. The equations of this module involve 15 parameters, of which 11 were considered as global (i.e. invariant over time and space) by the model's author, the remaining 4 being site-specific (Hénault et al., 2005). While the latter were laboratory-measured in all experimental sites and set to the resulting values throughout, the subset of 11 global parameters was estimated by our Bayesian procedure. We collated a database of N_2O flux measurements including 7 different field-sites in France, and various N fertilizer forms and rates in 2 of the sites. Bayesian calibration was applied either to each site or treatment individually, or directly to the ensemble of the data sets.

2.1 The CERES-EGC model

2.1.1 A process-based agro-ecosystem 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.

2.1.2 The nitrous oxide emission module

This module simulates the production of N_2O in soils through both the nitrification and the denitrification pathways, and was adapted from the semi-empirical model NOE (Hénault et al., 2005). The denitrification component is derived from the NEMIS model (Hénault and Germon, 2000) that calculates the actual denitrification rate (Da, kg N ha^{-1} d^{-1}) as the product of a potential rate at 20 °C (PDR, kg N ha^{-1} d^{-1}) with three unitless factors related to water-filled pore space (F_W), nitrate content (F_N) and temperature (F_T) in the topsoil, as follows:

$$Da = PDR\ F_N\ F_W\ F_T \tag{1}$$

In a similar fashion, the daily nitrification rate (Ni, kg N ha^{-1} d^{-1}) is modelled as the product of a maximum nitrification rate at 20 °C (MNR, kg N ha^{-1} d^{-1}) with three unitless factors related to water-filled pore space (N_W), ammonium concentration (N_N) and temperature (N_T) and expressed as follows:

$$Ni = MNR\ N_N\ N_W\ N_T \tag{2}$$

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

$$N_2O = r\ Da + c\ Ni \tag{3}$$

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. invariant over time and space. The local (site-specific) parameters are the potential denitrification rate (PDR), the maximum nitrification rate (MNR) and the fractions of nitrified (c) and denitrified (r) N that are evolved as N_2O. They were measured in the laboratory for all sites using a protocol that proved representative of field conditions in a wide range of situations (Hénault and Germon, 2000; Hénault et al., 2005; Gabrielle et al., 2006b; Dambreville et al., 2008). The 11 global parameters are the constants of the N_2O module equations which are considered invariant over time and space. They were estimated by Hénault and Germon (2000) for the denitrification pathway and by Garrido et al. (2002) and Laville et al. (2005) for nitrification. The equations of the response functions with the associated parameters are described in Appendix A (Eqs. 7-12). Prior information was gathered on all parameters on a literature review. For lack of information on the form of the pdf of these parameters, the latter were assigned uniform distributions within their likely range derived from literature data (Table 1). Parameters were supposed to be entirely independent (i.e. non-correlated). This type of hypotheses, which are likely to be violated in ecosystem models, is not a significant issue in the application of Bayesian calibration. For example, Naud et al. (2007) tested different levels of correlation of prior distributions and concluded that correlation was not a very important factor. In addition, Hong et al. (2005) reported that the assumption of a priori independence does not imply independence a posteriori, and the calibration may still provide a posterior estimate of correlations across parameters.

2.2 The database of N_2O measurements

The N_2O measurements were carried out on seven experimental sites located in Northern France. The experiments were conducted on major arable crop types and soils types representative of this part of France. For some sites, different treatments were conducted with various N-fertiliser amounts supplied to the crop, giving a total of 11 site/treatment combinations (Table 2). Nitrous oxide emissions were monitored by the static chamber method with eight replicates for all sites (Hénault et al., 2005), except at Grignon where measurements were monitored with three automatic chambers during 31 successive days from 13 May 2005 to 12 June 2005 (Lehuger et al., 2007). The variance in the measurements was estimated as the variance across the different replicate chambers in the field. Soil nitrogen and moisture contents were monitored in the soil profile for each site with different sampling frequencies (see references of Table 2 for details). The resulting samples were analysed for moisture content and inorganic N using colorimetric samples in the laboratory. Soil temperature was continuously monitored using thermocouples in most of the sites, except for the sites of Champnoël and Le Rheu. The input data required to run the model were also collected in each site: the weather data were taken from a local meteorological station, and detailed information on soil properties and crop management were compiled to generate CERES-EGC input files using a standard parameterization procedure (Gabrielle et al., 2006b). Uncertainty on these input data was not considered here since CERES-EGC had already been tested in most of the sites (Gabrielle et al., 2006b). Besides, it likely had little impact on the N_2O simulations since we checked that the model gave correct predictions of the major N_2O drivers (topsoil environmental conditions and nitrate content).

2.3 Bayesian calibration

2.3.1 Markov Chain Monte Carlo

Bayesian methods are used to estimate model parameters by combining two sources of information: prior information about parameter values and observations on output variables. The prior information is based on expert knowledge, literature review or by measuring parameters directly in the field or laboratory. In our case, the observations on output variables are field measurements of the different fluxes between soil-crop-atmosphere compartments. Bayes' theorem makes it possible to combine the two sources of information in order to calibrate the model parameters. The first step is to assign a probability distribution to the parameters, representing our prior uncertainty about their values. In our case, we specified lower and upper bounds of the parameters uncertainty, defining the prior parameter distributions as uniform. The aim of Bayesian calibration is to reduce this uncertainty by using the measured data, thereby producing the posterior distribution for the parameters. This is achieved by multiplying the prior with the likelihood function, which is the probability of the data given the parameters. The likelihood function is determined by the probability distribution of errors in observations. We assumed errors to be independent and normally distributed with mean zero following Van Oijen et al. (2005) and in the same fashion as Svensson et al. (2008) and Klemedtsson et al. (2007). Because probability densities may be very small numbers, rounding errors needed to be avoided and all calculations were carried out using logarithms. The logarithm of the data likelihood is thus set up, for each data set Y_i, as follows:

$$logL_i = \sum_{j=1}^{K} \left(-0.5 \left(\frac{y_j - f(\omega_i;\theta_i)}{\sigma_j} \right)^2 - 0.5log(2\pi) - log(\sigma_j) \right) \quad (4)$$

where y_j is the mean N_2O flux measured on sampling date j in the data set Y_i and σ_j the standard deviation across the replicates on that date, ω_i is the vector of model input data for the same date, $f(\omega_i;\theta_i)$ is the model simulation of y_j with the parameter vector θ_i, and K is the total number of observation dates in the data sets. To generate a representative sample of parameter vectors from the posterior distribution, we used a Markov Chain Monte Carlo (MCMC) method: the Metropolis-Hastings algorithm (Metropolis et al., 1953) (see Appendix B for details). We formed Markov chains of length 10^4-10^5 using a multivariate Gaussian pdf to generate candidate parameter vectors. The variance matrix of this Gaussian was tuned so that the Markov chains would explore parameter space efficiently. We followed the procedure of Van Oijen et al. (2005) and defined the variances equal to the square of 1 to 5 % of the prior parameter range (θ_{min}-θ_{max}) and zero covariances. Subsequently, the variances were tuned so that the fraction of candidates accepted during the random walk was between 20 to 30%. Ten percent of the total number of iterations at the beginning of the chain were discarded as unrepresentative "burn-in" of the chains (Van Oijen et al., 2005). For each calibration, three parallel Markov chains were started from three different starting points (θ_0): the default parameter value and their lower and upper bounds (θ_{min} and θ_{max}). Convergence was checked with the diagnostic proposed by Gelman and Rubin (1992), which is based on the comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance. Convergence is reached when variance between chains no longer exceeds the variance within each individual chain. The chains of parameter values resulting from the random walk of the Metropolis-Hastings algorithm are auto-correlated because each iteration depends on the previous one. We therefore thinned the chains 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 as the number of iterations between consecutive samples in a chain for which the auto-correlation was less than 60%. The chains filtered in this way were considered to be a representative sample from the posterior pdf, and from this sample were calculated the mean vector, the variance matrix and the 90% confident interval for each parameter.

The generation and analysis of the Markov chains were carried out with the statistical package R (R Development Core Team, 2008) and in particular its *coda* package (Plummer et al., 2006). The CERES-EGC model was encapsulated within R as a library, generated from the original Fortran code.

2.3.2 Procedure for the N_2O module

The calibration procedure had two main objectives: (i) to calibrate the parameters for each dataset Y_i, to explore the variations of global parameters across experimental sites and treatments, and (ii) to obtain better estimates for the global parameters (initially deemed universal in the model). The first objective was pursued by calibrating the parameters for each data set separately, which is referred to later on as the *dataset-by-dataset procedure*. In a second step, the global parameters were calibrated by running our procedure with the 11 data sets simultaneously (*multi-dataset procedure*), i.e. by calculating the posterior distribution as:

$$p(\theta|Y_1, ..., Y_{11}) \propto p(Y_1, ..., Y_{11}|\theta)\, p(\theta) \quad (5)$$

where Y_i is the data of the i^{th} site and the $\propto$ symbol means 'proportional to'. In this case, the log-likelihood is calculated as the sum of the log-likelihoods of all the data sets (for a given parameter set in the MCMC chain).

2.4 Evaluation of model predictions

The performance of the calibration procedures was assessed by calculating the root mean square error (RMSE). RMSE was defined, for each data set Y_i, as follows (Smith et al., 1996):

$$RMSE = \sqrt{\frac{\sum_{j=1}^{K}(y_j - f(\omega_i;\theta_i))^2}{K}} \quad (6)$$

In both following cases, simulations f($\omega_i; \theta_i$) were carried out using either the posterior expectancy of parameters ($\overline{\theta}$) or the maximum a posteriori (MAP) 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). In the case of prior parameter pdfs, the simulations were defined as the prior expectancy of the model predictions in which parameters were randomly drawn from the prior pdfs. For the posterior parameters pdfs, the simulations were the posterior expectancy of predictions. RMSE was computed after calibration resulting from the dataset-by-dataset or multi-dataset procedure.

3 Results

3.1 Simulation of soil state variables

Soil temperature, soil water content and nitrate and ammonium contents were simulated by the model and confronted against the measurements. Table 3 summarizes the mean deviation (MD), which is the mean difference between measurement and simulation, and RMSEs computed with the different topsoil state variables used as input variables of the N_2O emission module. Soil temperature and water content were well predicted by the model with RMSE ranging from 1.2 to 3.0 ° C for the soil temperature and from 3 to 6 % (v/v) for the soil water content across the 11 sites and treatments. The model's RMSE over the 11 sites and treatments ranged between 3.7 to 27.9 kg N ha^{-1} for the prediction of nitrate content and to 0.7 to 25.3 kg N ha^{-1} for the ammonium content. Dynamics of surface nitrate and ammonium contents were mainly driven by the fertiliser applications and mineralization of crop residues. Ammonium was rapidly nitrified across all the sites but the model failed to reproduce the background topsoil ammonium stock. Nitrate content was relatively well simulated except for 3 treatments for which N plant uptake was under-estimated (La Saussaye, Champnoël AN and Le Rheu AN).

3.2 Posterior parameter distributions

Figure 1 shows boxplots of the posterior parameter distributions after calibration with the dataset-by-dataset and the multi-dataset procedures. Such representation makes it possible to visualize differences between parameter pdfs across datasets, while the shape of the boxplot reveals the dispersion and symmetry of the marginal distributions. Our Bayesian procedure generally generated uni-modal distributions, and convergence test corroborated that the MCMC chains converged. Figure 2 presents the 50 and 97.5% quantiles of the Gelman-Rubin shrink factor for the 11 parameters calibrated with the data set of La Saussaye, and shows that it approached 1 for all parameters, evidencing the convergence of the calibration.

Figure 1 shows that the posterior distributions became narrower compared to the uniform prior distributions, which is undoubtedly due to the efficiency of our calibration procedure. The posterior pdfs converged to normal or log normal distributions, as already observed by Svensson et al. (2008) in the Bayesian calibration of a process-based forest model. Thus, the choice of an uniform distribution for the prior pdfs had little influence, as the information contained in the experimental data gradually became dominant in the calibration process (Van Oijen et al., 2005). For example, the posterior distributions of parameter θ_1 (the WFPS threshold triggering denitrification) had a narrow range for all datasets, suggesting that the calibration had drastically reduced its uncertainty. On the contrary, parameters θ_8 and θ_9 (corresponding to the minimum and maximum WFPS for nitrification activity, respectively) remained spread across their prior range of variation, and centered around their prior median. This means that the calibration did not significantly reduce their uncertainty. Conversely, some posterior distributions were flattened on one of the prior bounds, implying that their optimal values was outside the prescribed range. This was particular true for parameters θ_{10} (the half-saturation constant of nitrification response to ammonium) and θ_{11} (the Q10 factor for nitrification) for the data sets of Champnoël AN, La Saussaye and Grignon. We should therefore reconsider the prior ranges for these parameters.

The rightmost boxplot in each of the 11 graphs in Figure 1 depicts the distribution obtained with the multi-dataset procedure. The shape of this boxplot and its median value appeared to be more constrained by certain datasets than others, which may be explained by the fact that data sets with a comparatively larger number of observations of higher precision had substantially more weight in the log-likelihood function. For example, the boxplots of the multi-dataset calibration exhibited high similarity with those of the La Saussaye site for parameters θ_1, θ_3 and θ_6.

Some data sets were collected in the same sites, i.e. under identical climate patterns and soil types but with differentiated crop management (the Rafidin, Le Rheu and Champnoël datasets). Since the parameters of the N_2O module are mostly related to soil properties, it was expected that the calibration should produce similar distributions for these three sites. To a certain extent, this was the case for the parameters θ_2, θ_3 and θ_6, giving support to the idea that these parameters are mostly soil-dependent, and are little influenced by crop management. Conversely, the strong variation of posterior pdfs across sites challenges the original idea in model development that these parameters may be considered constant. The purpose of the multi-dataset procedure sought to investigate this option, by seeking the best-fit parameter pdfs in relation to the ensemble of the experimental situations collated in our database. It could be expected to lead to parameter pdfs with a wider spread (and thus higher uncertainty) than in the dataset-by-dataset calibration, owing to the wide ranges covered by the dataset-specific pdfs. While this was true of some parameters (e.g., θ_4, θ_5, and θ_7), it was the opposite for others (most notably θ_1 and θ_3).

Figure 3 depicts the ranges of response functions of the N_2O emission module resulting from the various calibrations, and evidences ample differences across datasets. The responses of nitrification to soil ammonium content (N_N, Fig. 3.a) were highly variable, reflecting the range taken by their shape parameter θ_{10}. The response of nitrification to soil WFPS (N_W, Fig. 3.b) shows that the minimum WFPS for nitrification activity (θ_8) were centred on a unique value, while the optimum WFPS (θ_7) was lower in the calibration with two data sets. The calibrated maximum

WFPSs for nitrification (θ_9) were centred on 90%. The shapes of the response function N_T (Fig. 3.c) were similar for two sites (La Saussaye and Grignon), but strikingly different for the other sites. The calibrated responses of denitrification to nitrate content (F_N, Fig. 3.d) were highly variable such as the response of nitrification to ammonium content. The shapes of the response of denitrification to WFPS (F_W) varied widely, as a consequence of the large variations of parameters θ_1 (the WFPS threshold triggering denitrification) and θ_6 (the exponent of the power-law). Hénault and Germon (2000) and Heinen (2006) showed that denitrification was highly sensitive to θ_1, and that this parameter was dependent on soil type. The response of denitrification to soil temperature (F_T) had a similar shape across the various parameterizations, for temperatures lower than 25 °C which corresponds to the range encountered in the field experiments. This leads to the conclusion that the function calibrated with the multi-dataset procedure could be considered universal.

Bayesian calibration also quantifies correlations between parameters in the posterior. Most parameters were cross-correlated, with coefficients higher than 0.4 for 6 of them (Table 1) suggesting that our uncertainty about their values is linked and implies that some parameters should be treated in clusters, as suggested by Svensson et al. (2008). Parameters θ_1 and θ_2 are positively correlated, and are both negatively correlated with θ_6.

3.3 Model prediction uncertainty

The simulations of N_2O emissions generated with the posterior MCMC parameter chains provided statistical distributions of model outputs resulting from parameter uncertainty, which is a straight benefit of Bayesian approaches. Figure 4 shows the mean of simulated daily N_2O emissions for all datasets (Fig. 4.a to 4.k). Some discrepancies between measurements and simulations remained, due to uncertainty on both sides. Measurement points with high standard deviations had less weight in the log likelihood function, and thus in the posterior probability, compared to lower fluxes with lower variability. For example, the two N_2O spikes measured in Villamblain in springtime (Fig. 4.a) had a large experimental error, but did not appear to constrain the calibration as much as the more frequent lower N_2O fluxes with much lower standard deviations. The same remark applies to Arrou (Fig. 4.b). For the dataset of Champnoël AN (Fig. 4.e), a high spike of N_2O was observed in autumn that the model failed to predict, whereas it otherwise successfully simulated fluxes under 10 g N_2O-N ha^{-1} d^{-1}.

For the Grignon site (Fig. 4.h), the observation points were concentrated on 31 successive days (from 13 May 2005 to 12 June 2005), and started a peak flux. With its default parameter set, the model simulated that peak along with two others in the following weeks that were not observed in the field (results not shown, see Lehuger et al. (2007)), in response to significant rains. The Bayesian calibration managed to circumvent the simulation of these two unobserved peak fluxes by raising the WFPS threshold for denitrification (θ_1) from 62% (default value) to 73%, which is the highest value in all the calibrations (Fig. 1.a). As a result of this change in the response to rainfall and soil water content, no N_2O-peaks were simulated throughout the year in Grignon (Fig. 4.h). For the dataset of Rafidin N0 (Fig. 4.i), observations also were concentrated on two short periods, but with fewer observations points than at Grignon. The calibration highly constrained the model during the measurement period, but appeared less constraining on the N_2O-fluxes outside this period.

Table 4 summarises the statistics of the annual N_2O emissions predicted by CERES-EGC for the different datasets. The mean annual fluxes ranged between 88 and 3672 g N_2O-N ha^{-1} y^{-1}, with a large confidence interval especially for the datasets with higher emission rates. An overall conversion factor of fertilizer inputs to N_2O-N was calculated as the ratio of the annual flux to the N fertiliser dose. This is different from an "emission factor", which takes background emissions of N_2O into account. Here, we also calculated this factor as the difference between the annual N_2O-N emissions of fertilised and unfertilised crops (g N_2O-N ha^{-1} y^{-1}) to the N-fertiliser dose.

The emission factors ranged from 0.05 and 1.12% across experimental sites, with a mean value of 0.26%. This value is four times lower than the default value recommended by the IPCC tier 1 methodology (IPCC, 2006).

3.4 Calibration efficiency and model prediction error

Table 5 summarises the RMSEs obtained with the various parameters sets, and made it possible to compare the efficiency of model calibration whether in the dataset-by-dataset or in the multi-dataset mode. In the dataset-by-dataset procedure, the RMSEs computed with the posterior expectancy of predictions were lower than those computed with the prior expectancy of predictions for all datasets except one (Arrou), with a 73% reduction on average and a maximum of 98% in La Saussaye. In 8 of the remaining 9 datasets, calibration lead to a reduction of 79% to 96% in the model's RMSE. On average across all datasets, the RMSE dropped from 39 down to 6 g N_2O-N ha^{-1} d^{-1} after calibration. There were no differences in the RMSEs calculated either with simulations based on the posterior mean of parameters ($\overline{\theta}$) or with posterior mean of predictions. Thus, the mean of our sample from the posterior could be directly used for the sites of our database or for sites with similar soil types. The use of the parameter set with maximum posterior probability (θ_{MAP}), i.e. when likelihood was maximum and given that we used a uniform prior, logically improved the RMSE compared to the use of the posterior mean of parameters ($\overline{\theta}$). As could be expected, the multi-dataset calibration was less efficient than the dataset-by-dataset one, enabling a decrease of only 33% of the RMSE computed with posterior expectancy of predictions compared to the prior expectancy of predictions. This would lead us to believe that the parameter set summarised in Table 1 could be a good compromise when the model will be applied for a new site.

In addition, Table 5 shows that the calibration did not really improve the simulations for two datasets: Villamblain and Arrou. For both datasets, the data were not informative enough to significantly improve parameter estimation. In the case of Arrou, the discrepancies may also be explained by the poor ability of CERES-EGC to simulate water-logging effects, as observed in this experiment. The N_2O module and in particular its denitrification part (Eqs. 1, 7, 8, 9 - Appendix A) were already shown unable of correctly rendering the dynamics of denitrification or N_2O emissions for soils with high degrees of water saturation. Still, RMSE values quantify the mismatch between simulations and the mean of the measurements without taking measurement uncertainty into account, or diagnosing whether problem lies with the simulations or the data. As a consequence, RMSE values should be interpreted with caution. More in-depth model evaluation would require comparing the behaviour of multiple models.

4 Discussion

4.1 Suitability and benefits of Bayesian calibration

Our main goal was to demonstrate the potential of a Bayesian-type calibration procedure to improve the parameterization of a N_2O-emission model, quantify parameter uncertainty and reduce uncertainties of model outputs. In recent years, Bayesian calibration was successfully applied to process-based ecosystem models, such as forest biomass growth models (Van Oijen et al., 2005; Svensson et al., 2008; Klemedtsson et al., 2007). Among the various possible Bayesian methods, MCMC is in principle particularly well adapted to such models (and in particular CERES-EGC) because they can handle a high number of parameters simultaneously (Makowski et al., 2002). Their efficiency is also not hampered by a poor knowlegde of the prior distributions, as is often the case with this type of models, and may be judged from the large variation range of the parameters we calibrated here. Method of expert elicitation have been recently developed and could be used in the future in order to refine prior distributions of model parameters. In short, elicitation is the process of translating expert knowledge about uncertain quantities into a probability distribution (Oakley and O'Hagan, 2007). However, no attempts had been made yet to calibrate processes so uncertain and irregular in time and space as N_2O emissions. This raised a number of issues in the adaptation of the MCMC algorithm. In particular, the chains were strongly auto-correlated, which required a substantial number of iterations (10^4 to 10^5), and drastic thinning. Also, the convergence had to be tested by running three parallel chains and using a variance-based diagnostic. An accurate simulation of the soil environmental drivers (temperature, moisture and mineral N contents) was a pre-requisite for the prediction of N_2O fluxes. Tests against field data showed that this condition was overall met, as noted in a previous test of CERES-EGC in a subset of the sites used here (Gabrielle et al., 2006b). In some instances, some discrepancies in the simulation of topsoil water content (Arrou) or nitrate content (La Saussaye, Champnoël AN and Le Rheu AN) which affected the prediction of N_2O fluxes. However, these errors point to structural deficiencies of the model (for instance in the simulation of soil water dynamics in the water-logged soil of Arrou), and did not interfere with the calibration. This was evidenced by the fact that inclusion of measured drivers improved model performance only marginally and in a few sites. This option was thus disregarded.

Our procedure significantly reduced parameter uncertainty for the datasets, and the uncertainty in simulated N_2O rates as a result. 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. 2, 10, 11, 12) of CERES-EGC (Rolland et al., 2008). The procedure we successfully implemented here may be readily used for other components of CERES-EGC, such as soil C turnover or crop photosynthesis and growth.

The calibration significantly reduced the model's RMSE compared with the prior parameter values, on average by 73% with the data-by-dataset procedure and by 33% with the multi-dataset procedure. Still, the calibration did not result in a perfect match between model simulations and observations of the 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 likelihood function. This may also be seen as an advantage since these extreme data points with large variance did not artificially influence the parameter values compared to lower-range values with better accuracy. Heinen (2006) also showed with a different calibration method that the optimised denitrification sub-module did not result in perfect fit at the daily compared to the seasonal scale.

Lastly, the dataset-by-dataset calibration points to ways of optimising calibration efficiency: when using manual chambers, N_2O measurements should be carried out at least once a month throughout the year, with a higher frequency during the peak fluxes subsequent to N-fertiliser and crop residues inputs and when soil conditions are favourable to denitrification , e.g. when soil moisture, soil temperature and mineralization rate are high.

4.2 Spatial variability of model parameters

We sought to calibrate model parameters either on a dataset-by-dataset basis in order to minimise model error or simultaneously on all datasets to find parameter values that would be universally applicable, following the premise behind the original development of the N_2O model. Such values would be extremely useful to apply the model to new soil conditions and to spatially extrapolate it. However, it was suggested that simple process-based models such as the one we used here needs to be parameterised on a site-specific basis (Heinen, 2006). The latter authors concluded to the impossibility of defining a set of response functions for denitrification (Eqs 1, 7 ,8, 9 - appendix A) that would equally apply to sandy, loamy and peat soil types. Our dataset-by-dataset calibration gave further evidence to that statement for the N_2O module of CERES-EGC, judging from the large variations in parameter pdfs across sites. However, our multi-dataset procedure also demonstrated that it is still possible to find global estimates for those parameters that encompass a wide range of experimental conditions, at the cost of a higher RMSE than with optimal, site-specific parameter sets. The parameter pdfs we obtained in the multi-dataset calibration shows which parameter values would be plausible, and may thus be used to improve the accuracy of N_2O simulations in new sites.

Models are often developed with the purpose of providing predictions over a large domain (in space and time). However, ensuring that their parameterisation is accurate is a pre-requisite to such application. When attempting at simulating N_2O fluxes in a new site where no measured data are available, the results of our calibration points to the following strategy to meet this requirement. First, the user should check if calibrated parameter sets already exist for similar soil types, based on soil taxonomy or physico-chemical characteristics. If not, the parameter values derived from the multi-dataset calibration may be used. They may also serve as default values for the spatial extrapolation of the model at the regional scale. In the future, new data sets may be assimilated in the calibration to reduce the uncertainty of global parameters and to increase the application domain of the model. Alternatively, it is clearly advisable to favour the collection of N_2O emissions data for the new sites, which lead to a much better performance of the model. One last obstacle to the extrapolation of CERES-EGC lies in the 4 site-specific parameters, which are supposed to be measured in the laboratory. We chose to exclude them from the calibration in accordance with the original model design. However, including them would be interesting to simulate a situation where such experimental determination is not possible, and to see to what extent it influences the outcome of the calibration. It is likely to result in different parameter values since, for instance, the potential denitrification rate (a local parameter) was shown to significantly correlate with three global parameters related to denitrification (Gabrielle, 2006). However, testing such a scenario appeared beyond the scope of this paper since it implied too strong a deviation from the model hypotheses.

4.3 Prediction of N_2O fluxes from agro-ecosystems

CERES-EGC and its specific N_2O module have already been used in a range of soil conditions (Hénault et al., 2005; Dambreville et al., 2008; Heinen, 2006), and model uncertainty had only been quantified using simple Monte Carlo techniques for a subset of 5 parameters (Gabrielle et al., 2006a). The effect of parameter uncertainty was seldom analysed with ecosystem models simulating N_2O emissions, although (or perhaps also because) N_2O measurements are fraught with a daunting spatial and temporal variability (Duxbury and Bouldin, 1982). Our Bayesian calibration resulted in a probabilistic simulation of the time course of N_2O emissions taking such variability and uncertainty into account, through their consequences on parameters' distributions. The calibrated model could predict daily N_2O fluxes rather well, except for the highest peaks with high experimental error which it failed to predict in some cases.

In addition, the procedure makes it possible to quantify model output uncertainty in the calculation of annual N_2O budget and emission factors (EFs). The model predicted annual N_2O fluxes were ranging from 88 to 3672 g N_2O-N $ha^{-1}y^{-1}$ over the various sites, and EFs ranging from 0.05 to 1.12%. On the basis of these results, alongside those of Gabrielle et al. (2006a), it appears that the 1% default EF value of the IPCC Tier 1 methodology is not suitable for the sites we studied because it would considerably overestimate the annual emissions (Table 4). In Belgium, Beheydt et al. (2007) used the DNDC model to calculate EFs corresponding to various scenarios involving high N input levels and N surpluses, and obtained an average value of 6.49%, which is 25 times higher than ours, compared to an estimate of 3.16% using the N_2O measurements. Their observed emission range was an order of magnitude higher than that of our database. Assimilate such extreme data with our procedure would be helpful to enlarge the prediction range of CERES-EGC, and to check its ability to predict annual emissions higher than 10 kg N_2O-N ha^{-1} y^{-1}.

Our results also suggested that annual N_2O emissions were not strictly proportional to fertiliser

N rate, which is in agreement with the results of Barton et al. (2008). The latter 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. They concluded that the emissions of N_2O from arable soils could not be directly derived from the application of N fertiliser, and that other factors (e.g., soil properties) should be taken into account.

Bayesian calibration provided valuable insight into the uncertainty of the simulated N_2O fluxes, making it possible to take risk into account in a range of model applications: estimation of the global warming potential (GWP) of agro-ecosystems, assessment of cropping systems' environmental balance, or decision support in agriculture. It would also be interesting to compare the ability of various agro-ecosystem models to predict N_2O emissions on the same data sets, in a similar fashion as Frolking et al. (1998) and 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. Finally, the outputs of several models could be combined to improve the accuracy of the prediction, as was suggested with atmospheric models (Fisher et al., 2002).

5 Conclusion and future work

Bayesian calibration was successfully applied to the CERES-EGC agro-ecosystem model to improve the parameterization of its N_2O emission module, thanks to a careful analysis and diagnostic of the MCMC chains of parameters generated by the Metropolis-Hastings algorithm. The parameters were calibrated either (i) against separately data sets or (ii) by using all the data sets simultaneously, to satisfy our objectives which were, respectively, to improve model simulations at the field scale and to find universal values of parameters in order to spatially extrapolate the model. In addition, Bayesian calibration provided a means of quantifying uncertainties in both parameters and model outputs. 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 multi-dataset 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.

A remaining obstacle to the extrapolation of the N_2O module lies in the 4 local parameters that should be measured or estimated on site (Hénault et al., 2005), and that were accordingly not calibrated here. Identifying the key soil or landscape characteristics that control these parameters appears as a pre-requisite to the large-scale use of CERES-EGC.

Based on our results, we recommended a strategy to deal with model extrapolation and parameters' variability. Nevertheless, another option to tackle spatial variability would consist in using other types of prior information (e.g. on soil properties) to infer the parameters of the N_2O module. In future work, it would be beneficial to identify such "hyperparameters" which may explain spatial variability (Clark, 2005), and to develop a hierarchical Bayesian approach to derive their pdfs.

Acknowledgements

This work was part of the NitroEurope Integrated Project (EU's Sixth Framework Programme for Research and Technological Development) which investigates the nitrogen cycle and its influence on the European greenhouse gas balance. We wish to thank Matieyiendu Lamboni and Hervé Monod (INRA Jouy-en-Josas) for useful advice and discussions. Special thanks to Christophe Dambreville for making available the data from the Champnoël and Le Rheu sites.

Appendix A. Equations of the nitrous oxide emission module

The response functions are unitless and read:

$$F_N = \frac{\left[NO_3^-\right]}{Km_{denit} + \left[NO_3^-\right]} \tag{7}$$

where F_N is the denitrification response factor to $[NO_3^-]$ the soil nitrate content (mg N kg^{-1} soil), and Km_{denit} the half-saturation constant (mg N kg^{-1} soil).

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

$$F_W = \left[\frac{WFPS - Tr_{WFPS}}{1 - Tr_{WFPS}}\right]^{POW},\ WFPS \geq Tr_{WFPS} \tag{8}$$

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 law.

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

$$F_T = exp\left[\frac{(T-20)ln(Q10_{denit,2})}{10}\right],\ T \geq TTr_{denit} \tag{9}$$

where F_T is the denitrification response function to soil temperature (T, °C), in the form of two sequential Q10 functions below and above a threshold temperature (TTr_{denit}). The two $Q_1 0$ values ($Q10_{denit,1}$ and $Q10_{denit,2}$) correspond to the relative increase in denitrification activity for every 10 °C increase in T.

$$N_N = \frac{\left[NH_4^+\right]}{Km_{nit} * Hp + \left[NH_4^+\right]} \tag{10}$$

where N_N is the nitrification response factor to $[NH_4^+]$, the soil ammonium content (mg N kg^{-1} soil). The half-saturation constant Km_{nit} (mg N kg^{-1} soil) is calculated at each soil water content (Hp, w/w).

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

$$N_W = \frac{MAX_{WFPS} - WFPS}{MAX_{WFPS} - OPT_{WFPS}},\ OPT_{WFPS} \leq WFPS < MAX_{WFPS} \tag{11}$$

$$else\ N_W = 0$$

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).

$$N_T = exp\left[\frac{(T-20)ln(Q10_{nit})}{10}\right] \tag{12}$$

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

Appendix B. The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm consists of three steps:

Step 1. Randomly generate a new "candidate" parameter vector

$$\theta^* = \theta_{i-1} + \delta \tag{13}$$

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

Step 2. Calculate the ratio of the posterior probability of the candidate vector over the posterior probability of the current candidate:

$$\alpha = \frac{p(\theta^*|Y)}{p(\theta_{i-1}|Y)} = \frac{p(Y|\theta^*)p(\theta^*)}{p(Y|\theta_{i-1})p(\theta_{i-1})} \tag{14}$$

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 the candidate vector minus the log of the posterior probability of the current vector.

Step 3. Accept θ^* if $\alpha \geq u$ 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.

References

Barton, L., Kiese, R., Gatter, D., Butterbach-Bahl, K., Buck, R., Hinz, C., Murphy, D. V., 2008. Nitrous oxide emissions from a cropped soil in a semi-arid climate. Glob. Change Biol. 14, 177–192.

Bateman, E. J., Baggs, E. M., 2005. Contributions of nitrification and denitrification to N_2O emissions from soils at different water-filled pore space. Biol. Fertil. Soils 41, 379–388.

Beheydt, D., Boeckx, P., Sleutel, S., Li, C. S., Van Cleemput, O., 2007. Validation of DNDC for 22 long-term N_2O field emission measurements. Atmos. Environ. 41, 6196–6211.

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

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

Clark, J. S., 2005. Why environmental scientists are becoming bayesians. Ecol. Lett. 8, 2–14.

Dambreville, C., Morvan, T., Germon, J. C., 2008. N_2O emission in maize-crops fertilized with pig slurry, matured pig manure or ammonium nitrate in Brittany. Agric. Ecosys. Environ. 123, 201–210.

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

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Walsh, M. K., Ojima, D. S., Thornton, P. E., 2006. DAYCENT national-scale simulations of nitrous oxide emissions from cropped soils in the united states. J. of Environ. Qual. 35, 1451–1460.

Ding, W. X., Cai, Y., Cai, Z. C., Yagi, K., Zheng, X. H., 2007. Nitrous oxide emissions from an intensively cultivated maize-wheat rotation soil in the North China Plain. Sci. Total Environ. 373, 501–511.

Dobbie, K. E., Smith, K. A., 2001. The effects of temperature, water-filled pore space and land use on N_2O emissions from an imperfectly drained gleysol. Eur. J. Soil Sci. 52, 667–673.

Duxbury, J. M., Bouldin, D. R., 1982. Emission of nitrous oxide from soils. Nature. 298, 462–464.

Fisher, B. E. A., Ireland, M. P., Boyland, D. T., Critten, S. P., 2002. Why use one model? An approach for encompassing model uncertainty and improving best practice. Environ. Model. and Assess. 7, 291–299.

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., Smith, K. A., 1998. Comparison of N_2O emissions from soils at three temperate agricultural sites: simulations of year-round measurements by four models. Nutr. Cycl. Agroecosys. 52, 77–105.

Gabrielle, B., 2006. Sensitivity and uncertainty analysis of a static denitrification model. In Wallach, D., Makowski, D., 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., Hénault, C., 2006a. Process-based modeling of nitrous oxide emissions from wheat-cropped soils at the subregional scale. Global Biogeochem. Cycles. 20. doi:10.1029/2006GB002686.

Gabrielle, B., Laville, P., Hénault, C., Nicoullaud, B., Germon, J. C., 2006b. Simulation of nitrous oxide emissions from wheat-cropped soils using CERES. Nutr. Cycl. Agroecosys. 74, 133–146.

Gabrielle, B., Menasseri, S., Houot, S., 1995. Analysis and field-evaluation of the CERES models water-balance component. Soil Sci. Soc. Am. J. 59, 1403–1412.

Gallagher, M., Doherty, J., 2007. Parameter estimation and uncertainty analysis for a watershed model. Environ. Model. Software 22, 1000–1020.

Garrido, F., Hénault, C., Gaillard, H., Perez, S., Germon, J. C., 2002. N_2O and NO emissions by agricultural soils with low hydraulic potentials. Soil Biol. Biochem. 34, 559–575.

Gelman, A., Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 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., 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, 119–124.

Grant, R. F., Pattey, E., 2003. Modelling variability in N_2O emissions from fertilized agricultural fields. Soil Biol. Biochem. 35, 225–243.

Heinen, M., 2006. Application of a widely used denitrification model to Dutch data sets. Geoderma. 133, 464–473.

Hénault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J. C., Cellier, P., 2005. Predicting in situ soil N_2O emission using NOE algorithm and soil database. Glob. Change Biol. 11, 115–127.

Hénault, C., Germon, J. C., 2000. NEMIS, a predictive model of denitrification on the field scale. Eur. J. Soil Sci. 51, 257–270.

Hong, B. G., Strawderman, R. L., Swaney, D. P., 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 Resourc. Res. 41, doi:10.1029/2004WR003551.

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., Delmas, R., 1997. Quantification of N-losses as NH_3, NO, N_2O and N_2 from fertilized maize fields in Southwestern France. Nutr. Cycl. Agroecosys. 48, 91–104.

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

Jones, C. A., 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., Raftery, A. E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Klemedtsson, L., Jansson, P.-E., Gustafsson, D., Karlberg, L., Weslien, P., von Arnold, K., Ernfors, M., Langvall, O., Lindroth, A., 2007. Bayesian calibration method used to elucidate carbon turnover in forest on drained organic soil. Biogeochemistry.doi:10.1007/s10533-007-9169-0.

Larssen, T., Huseby, R. B., Cosby, B. J., Host, G., Hogasen, T., Aldrin, M., 2006. Forecasting acidification effects using a Bayesian calibration and uncertainty propagation approach. Environ. Sci. Technol. 40, 7841–7847.

Laville, P., Hénault, C., Gabrielle, B., Serca, D., 2005. Measurement and modelling of NO fluxes on maize and wheat crops during their growing seasons: effect of crop management. Nutr. Cycl. Agroecosys. 72, 159–171.

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

Li, C., Zhuang, Y., Cao, M., Crill, P., Dai, Z., Frolking, S., Moore III, B., Salas, W., Song, W., Wang, X., 2001. Comparing a process-based agro-ecosystem model to the IPCC methodology for developing a national inventory of N_2O emissions from arable lands in China. Nutr. Cycl. Agroecosys. 60, 159–175.

Li, C. S., 2000. Modeling trace gas emissions from agricultural ecosystems. Nutr. Cycl. Agroecosys. 58, 259–276.

Li, Y., Chen, D. L., Zhang, Y. M., Edis, R., Ding, H., 2005. Comparison of three modeling approaches for simulating denitrification and nitrous oxide emissions from loam-textured arable soils. Global Biogeochem. Cycles. 19. doi:10.1029/2004GB002392.

Linn, D. M., 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, 1267–1272.

Lokupitiya, E., Paustian, K., 2006. Agricultural soil greenhouse gas emissions: A review of National Inventory Methods. J. Environ. Qual. 35, 1413–1427.

Maag, M., Vinther, F. P., 1996. Nitrous oxide emission by nitrification and denitrification in different soil types and at different soil moisture contents and temperatures. Appl. Soil Ecol. 4, 5–14.

Maag, M., Vinther, F. P., 1999. Effect of temperature and water on gaseous emissions from soils treated with animal slurry. Soil Sci. Soc. Am. J. 63, 858–865.

Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M. H., 2006. Parameter estimation for crop models. In Wallach, D., Makowski, D., Jones, J. W., editors, Working with dynamic crop models: evaluating, analyzing, parameterizing and using them. Chapter 4. Elsevier.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Monod, H., Naud, C., Makowski, D., 2006. Uncertainty and sensitivity analysis for crop models. In Wallach, D., Makowski, D., and Jones, J. W., editors, Working with dynamic crop models: evaluating, analyzing, parameterizing and using them. Chapter 3. Elsevier.

Naud, C., Makowski, D., and Jeuffroy, M. H., 2007. Application of an interacting particle filter to improve nitrogen nutrition index predictions for winter wheat. Ecol. Model. 207, 251–263.

Oakley, J. E., and O'Hagan, A., 2007. Uncertainty in prior elicitations: a nonparametric approach. Biometrika 94: 427–441.

Parton, W. J., Holland, E. A., Del Grosso, S. J., Hartman, M. D., Martin, R. E., Mosier, A. R., Ojima, D. S., Schimel, D. S., 2001. Generalized model for NO_x and N_2O emissions from soils. J. Geophys. Res-Atmos. 106, 17403–17419.

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

Pattey, E., Edwards, G. C., Desjardins, R. L., Pennock, D. J., Smith, W., Grant, B., MacPherson, J. I., 2007. Tools for quantifying N_2O emissions from agroecosystems. Agr. Forest Meteorol. 142, 103–119.

Pihlatie, M., Syvasalo, E., Simojoki, A., Esala, M., Regina, K., 2004. Contribution of nitrification and denitrification to N_2O production in peat, clay and loamy sand soils under different soil moisture conditions. Nutr. Cycl. Agroecosys. 70, 135–141.

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

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., Stengel, P., 1994. Oxygen-transport and anaerobiosis in aggregated soils - contribution to the study of denitrification. Agronomie. 14, 395–409.

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

Rolland, M. N., Gabrielle, B., Laville, P., Serca, D., Cortinovis, J., Larmanou, E., Lehuger, S., Cellier, P., 2008. Modeling of nitric oxide emissions from temperate agricultural soils. Nutr. Cycl. Agroecosys. 80, 75–93.

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

Smith, J., Smith, P., Addiscott, T., 1996. Quantitative methods to evaluate and compare soil organic matter (SOM) models. In Evaluation of soil organic matter models using existing long-term datasets (eds Powlson, D., Smith, P., and Smith, J. E.), 183–202. NATO ASI Series 1, Vol. 38. Springer-Verlag, Heidelberg.

Smith, K. A., 1997. The potential for feedback effects induced by global warming on emissions of nitrous oxide by soils. Glob. Change Biol. 3, 327–338.

Smith, K. A., Thomson, P. E., Clayton, H., McTaggart, I. P., Conen, F., 1998. Effects of temperature, water content and nitrogen fertilisation on emissions of nitrous oxide by soils. Atmos. Environ. 32, 3301–3309.

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., OMara, F., Rice, C., Scholes, B., 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., Vander Pol, R. A., 1975. Effect of Temperature on Denitrification Rate in Soils. Soil Sci. Soc. Am. J. 39, 867–870.

Svensson, M., Jansson, P. E., Gustafsson, D., Kleja, D. B., Langvall, O., Lindroth, A., 2008. Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand. Ecol. Model. 213, 331–344.

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

List of Tables

Parameter vector $\theta = [\theta_1 ... \theta_{11}]$					Prior probability distribution			Posterior probability distribution		
θ_i	Symbol	Description	Unit	Default value	θ_{min}(i)	θ_{max}(i)	References	Mean	SD	Correlated $\{\theta_i\}$
θ_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	$\{2,\underline{6}\}$
θ_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	$\{1,\underline{6}\}$
θ_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	
θ_4	$\mathrm{Q10}_{denit,1}$	Q10 factor for low temperature	Unitless	89.00	60.00	120.00	Stanford et al. (1975); Maag and Vinther (1999)	89.46	18.28	$\{\underline{5}\}$
θ_5	$\mathrm{Q10}_{denit,2}$	Q10 factor for high temperature	Unitless	2.10	1.00	4.80	Gabrielle (2006); Stanford et al. (1975)	2.62	1.17	$\{\underline{4},10\}$
θ_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	$\{\underline{1}, \underline{2}\}$
θ_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	
θ_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	
θ_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	
θ_{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	$\{5\}$
θ_{11}	$\mathrm{Q10}_{nit}$	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 a literature review. The posterior parameter distributions are based on the multi-dataset procedure, and are characterised by the mean value of the posterior, their standard deviation (SD). Correlations with other parameters are reported if their absolute value exceeds 0.4 (underlined parameters express a negative correlation).

Site	Treatment	Year	Soil texture class	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: Main characteristics of the N_2O emissions data base used in the model calibration. At Rafidin, the treatments N0, N1 and N2 correspond to various N-fertilizer applications and at Le Rheu and Champnoël, the treatments AN correspond to ammonium nitrate application and CT to the control plot.

Site	Treatment	Soil temperature				Soil water content				Nitrate content				Ammonium content			
		N	Mean	MD	RMSE	N	Mean	MD	RMSE	N	Mean	MD	RMSE	N	Mean	MD	RMSE
			(°C)				(v/v)				(kg NO_3-N ha^{-1})				(kg NH_4-N ha^{-1})		
Rafidin	N0	294	8.7	-1.2	3.0	20	0.253	-0.027	0.043	21	10.8	5.5	9.9	21	3.7	3.5	4.1
	N1	294	8.7	-1.2	3.0	20	0.244	-0.035	0.051	21	12.9	8.0	11.8	21	5.6	5.0	6.8
	N2	294	8.7	-1.2	3.0	20	0.240	-0.039	0.050	21	23.5	17.0	22.6	21	6.2	5.6	8.0
Villamblain		250	8.4	0.1	1.3	7	0.344	0.024	0.027	7	17.6	8.7	11.0	7	6.5	4.8	6.0
Arrou		250	8.4	0.2	1.2	7	0.343	0.053	0.056	7	18.1	11.8	14.9	7	9.1	9.0	10.6
La Saussaye		250	8.4	-1.2	2.4	7	0.307	0.030	0.038	7	15.3	-15.9	27.9	7	5.9	5.8	8.8
Champnoël	CT	no data	no data	no data	no data	14	0.239	-0.009	0.049	2	28.8	3.4	3.7	2	0.9	0.7	0.7
	AN	no data	no data	no data	no data	14	0.239	-0.006	0.027	11	22.4	-20.8	29.5	11	13.4	8.0	14.6
Le Rheu	CT	no data	no data	no data	no data	13	0.212	0.004	0.028	9	17.8	-1.4	15.4	9	4.6	4.3	4.6
	AN	no data	no data	no data	no data	13	0.212	0.004	0.028	10	54.3	-16.6	27.8	10	4.5	-7.8	25.3
Grignon		364	11.7	-0.4	2.4	13	0.249	0.002	0.028	11	71.4	-1.7	14.2	11	12.3	6.8	13.6

Table 3: Sample size (N), mean of measured in situ soil variables (Mean), mean deviation (MD) and root mean square errors (RMSE) computed with the predicted and measured soil variables: soil temperature, soil water content and topsoil nitrate and ammonium contents for the 11 data sets.

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: Annual N_2O fluxes (g N_2O-N $ha^{-1}y^{-1}$) calculated as the sum of mean, 0.05 and 0.95 quantiles of daily simulations with the calibrated parameter sets. Annual estimates from IPCC methodology (corresponding to the emissions due to fertiliser application), conversion factor (%) and emission factor (%) are also reported (see text for definition), along with their 90% confidence band.

Site	Treatment	RMSE (in g N_2O-N ha^{-1} d^{-1}) computed with:				
		Prior expectancy of predictions	Posterior expectancy of predictions	Posterior expectancy of parameters	Maximum a posteriori parameter vector	Posterior expectancy of predictions with the multi-dataset procedure
Rafidin	N0	4.6	0.7	0.3	0.3	4.6
	N1	7.5	1.2	1.4	1.2	12.8
	N2	10.5	2.1	3.0	2.8	20.4
Villamblain		5.2	4.8	4.9	4.9	5.5
Arrou		25.4	27.1	25.3	23.8	29.2
La Saussaye		93.0	2.0	2.3	2.4	2.3
Champnoël	CT	21.5	1.4	0.9	0.9	0.9
	AN	65.58	13.8	14.0	13.8	14.0
Le Rheu	CT	149.5	6.1	6.0	6.0	6.0
	AN	30.4	2.0	2.2	2.2	2.4
Grignon		16.9	1.0	1.2	1.3	1.1

Table 5: Root mean square errors (RMSE, in g N_2O-N ha^{-1} d^{-1}) based on: the prior expectancy of predictions, the posterior expectancy of predictions, the posterior expectancy of parameters, the maximum a posteriori parameter vector and the posterior expectancy of predictions from the multi-dataset procedure.

List of Figures

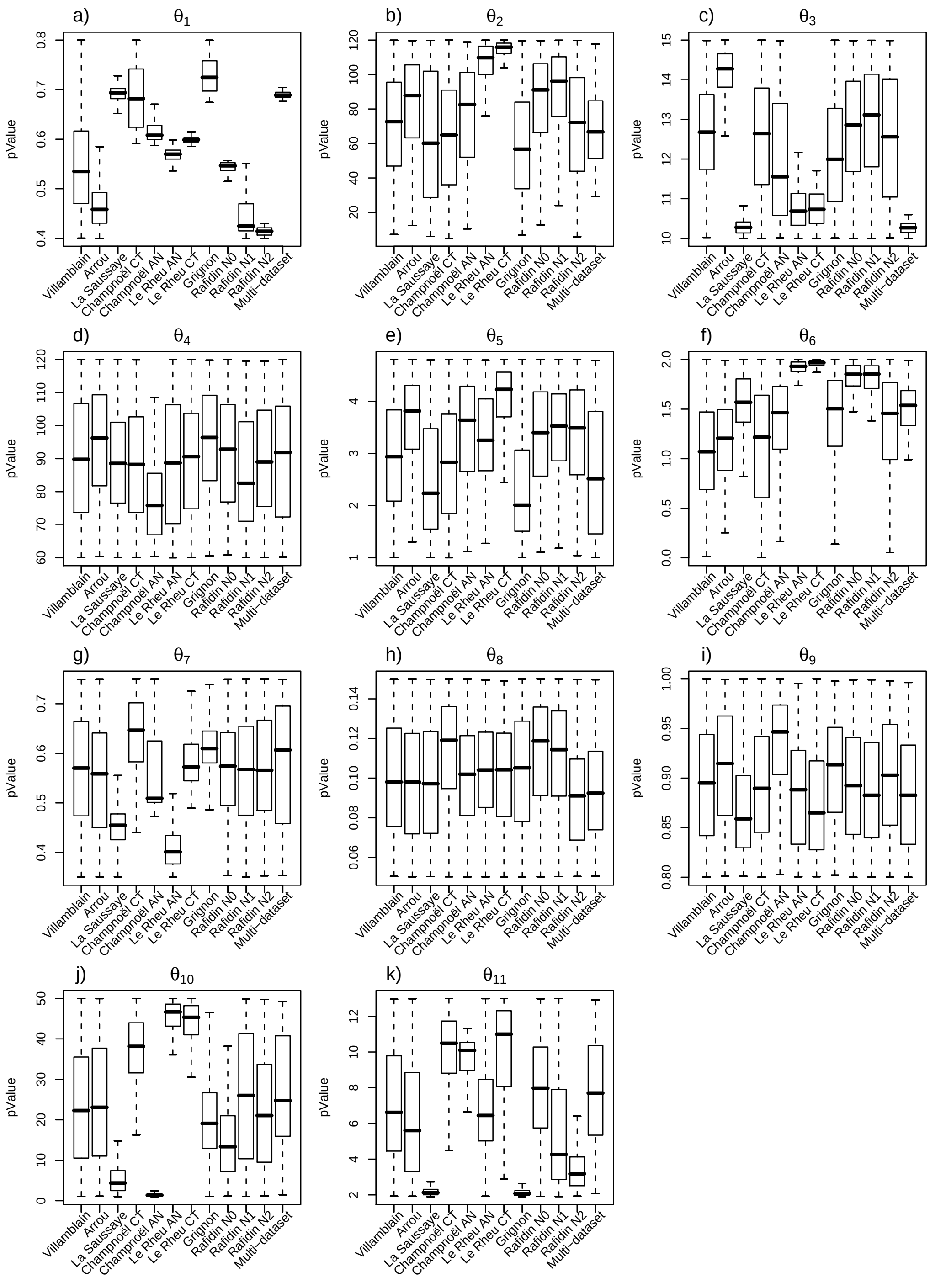

Figure 1: Posterior distributions of the 11 calibrated parameters (θ_1 to θ_{11}) represented as boxplots over the prior range of variation (corresponding to the range of the y-axis). The boxplots are computed from calibration dataset-by-dataset and with the "multi-dataset" procedure. The boxplots depict the median (solid line), the 2nd and 3rd quartiles (bars), the 1st and 4ht quartiles (dotted line), and the extreme values (excluding outliers).

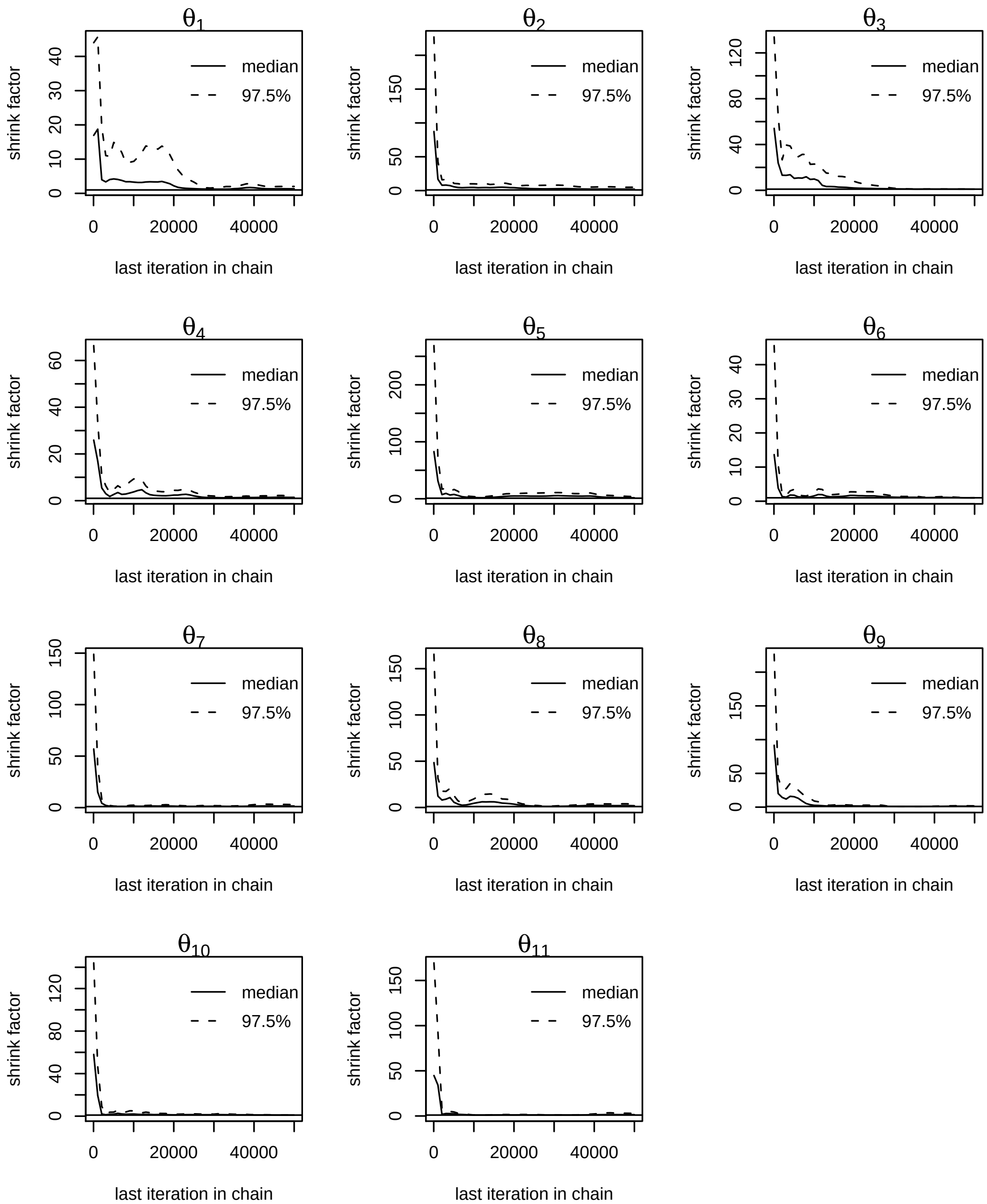

Figure 2: Evolution of the Gelman and Rubin's shrink factor for the calibration of the site La Saussaye.

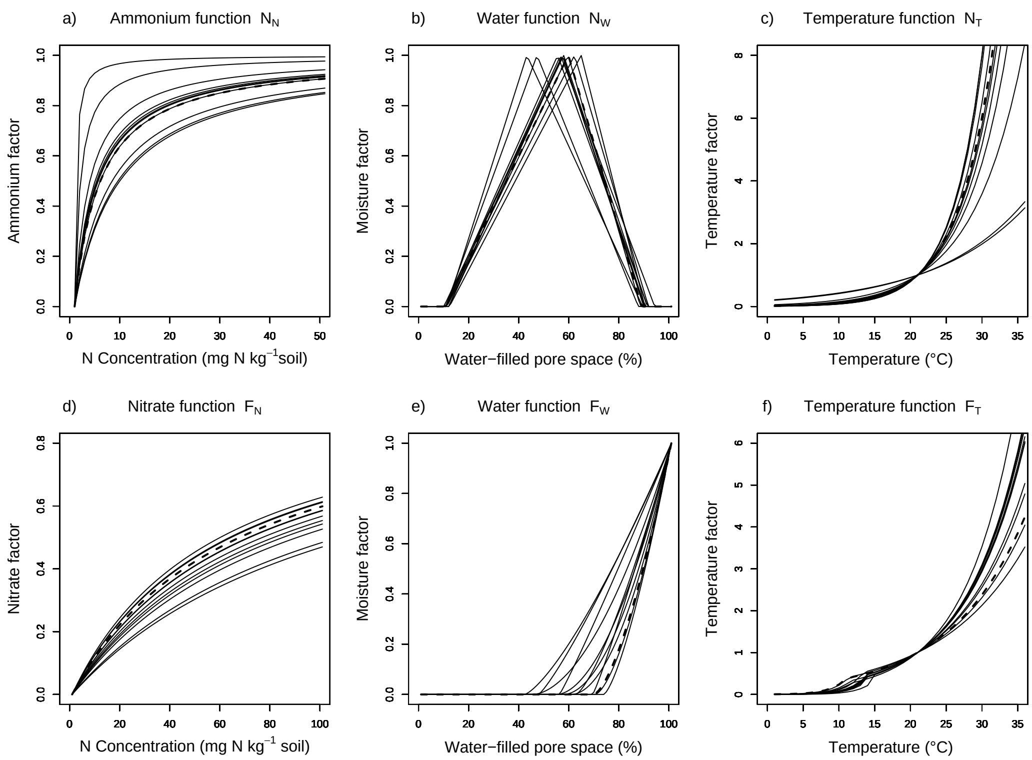

Figure 3: Response functions of the N_2O emission module traced with different parameters sets: mean of the posterior for each dataset-by-dataset calibration (line), and mean of the posterior for the multi-dataset calibration (dashed line).

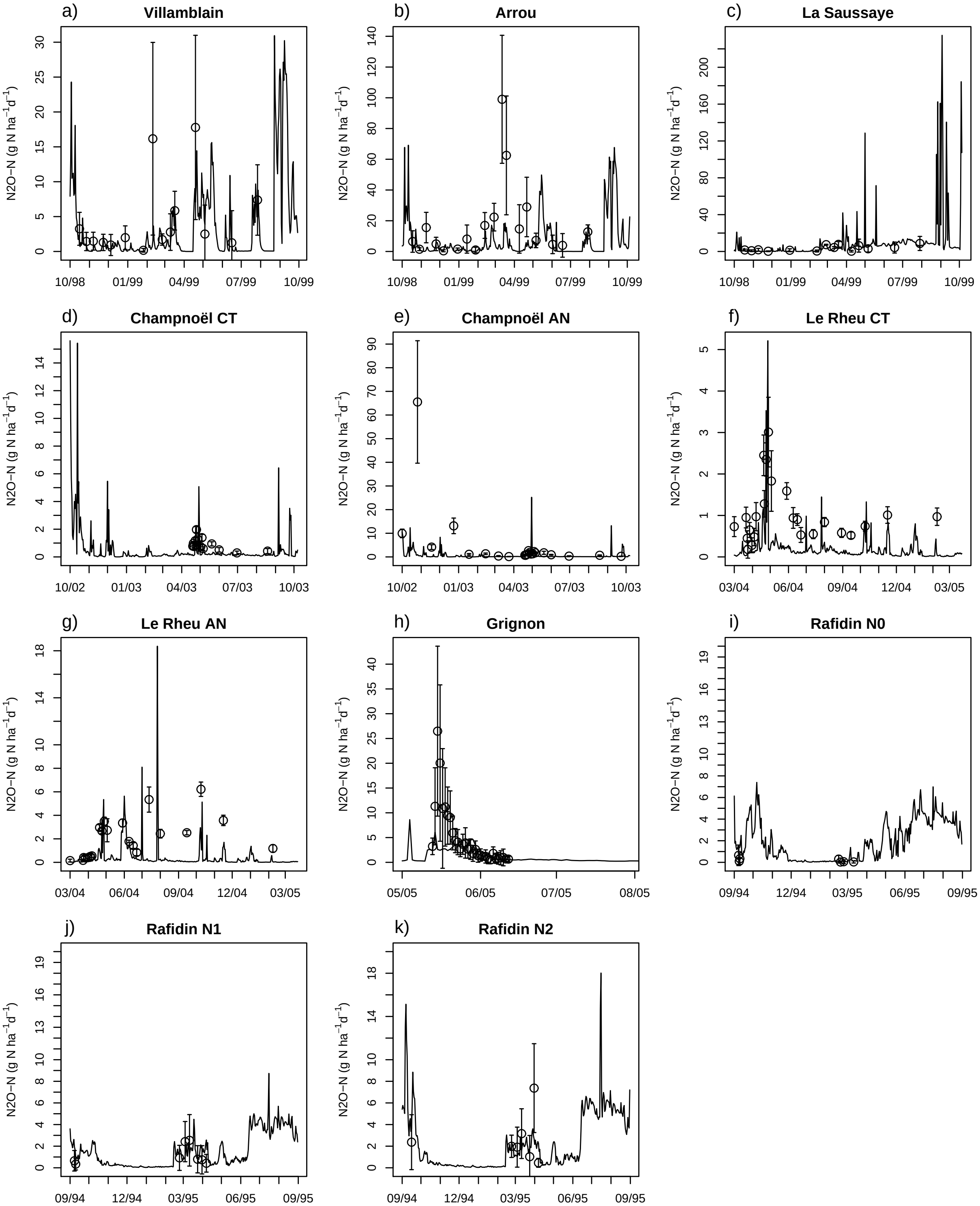

Figure 4: Simulated (lines) and observed (symbols) N_2O emissions for the different sites and treatments. The simulated line is the posterior expectancy of predictions from dataset-by-dataset calibrations.