# Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods

**Article** *in* Agronomie · March 2002
DOI: 10.1051/agro:2002007



**3 authors:**

David Makowski
French National Institute for Agricultural Research
**241** PUBLICATIONS **3,715** CITATIONS
SEE PROFILE

Daniel Wallach
French National Institute for Agricultural Research
**180** PUBLICATIONS **4,312** CITATIONS
SEE PROFILE

Marie Tremblay
L'Institut national de la recherche agronomique (Toulouse)
**42** PUBLICATIONS **570** CITATIONS
SEE PROFILE

Agronomie 22 (2002) 191–203

DOI: 10.1051/agro:2002007



**Original article**

# Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods

David MAKOWSKI[a]*, Daniel WALLACH[b], Marie TREMBLAY[b]

[a] Unité d'Agronomie, INRA Grignon, 78850 Thiverval-Grignon, France
[b] Unité d'Agronomie, INRA Toulouse, BP 27, 31326 Castanet-Tolosan, France

(Received 1 March 2001; revised and accepted 15 January 2002)

**Abstract –** The Bayesian approach allows one to estimate model parameters from prior expert knowledge about parameter values and from experimental data. The purpose of this paper is to compare the performances of two Bayesian methods, namely the Metropolis-Hastings algorithm and the Generalized Likelihood Uncertainty Estimation method (GLUE). These two methods are applied to a non-linear model that includes 22 parameters. This model has the main features of an agronomic model. The two Bayesian methods give similar results. The parameter estimates obtained with the two methods have similar properties. Both methods improve strongly the accuracy of model predictions even when only few data samples are available for estimating the parameters. However, the values of mean squared error of prediction of the model are slightly higher with the GLUE method than with the Metropolis-Hastings algorithm. The performances of the methods are sensitive to the prior assumptions made about parameter values.

**Bayes / Markov chain Monte Carlo / parameter estimation / parameter uncertainty**

**Résumé – Une approche Bayésienne pour estimer les paramètres des modèles de cultures, comparaison des méthodes GLUE et MCMC.** L'approche Bayésienne permet d'estimer les paramètres des modèles en prenant en compte à la fois les connaissances a priori des experts sur les valeurs des paramètres et les données expérimentales. Le but de cet article est de comparer les performances de deux méthodes Bayésiennes : l'algorithme de Metropolis-Hastings et la méthode GLUE. Ces deux méthodes sont appliquées à un modèle non linéaire incluant 22 paramètres. Ce modèle a les principales caractéristiques des modèles agronomiques. Les deux méthodes donnent des résultats voisins. Les estimateurs des paramètres obtenus avec les deux méthodes sont de qualités comparables. Les deux méthodes améliorent la précision des prédictions du modèle, même lorsque peu de données sont disponibles pour estimer les paramètres. Cependant, les valeurs du mean squared error of prediction obtenues avec la méthode GLUE sont légèrement supérieures à celles obtenues avec l'algorithme de Metropolis-Hastings. Les performances des méthodes sont par ailleurs très dépendantes des hypothèses faites a priori sur les valeurs des paramètres.

**Bayes / estimation des paramètres / Markov chain Monte Carlo / incertitude des paramètres**

**Communicated by Gérard Guyot (Avignon, France)**

* Correspondence and reprints
makowski@grignon.inra.fr

## 1. INTRODUCTION

A proper estimation of model parameters is required for ensuring accurate model predictions and good model-based decision rules. Two types of information are generally available for estimating the parameters of agronomic models: prior information based on expert knowledge and data collected via experimentation. With most statistical methods, such as the least squares method, only the data are used for estimating model parameters. Such methods require users to blind themselves to other existing information. Contrary to these methods, the Bayesian approach allows one to take into account not only data but also prior information about parameter values for estimating model parameters. Another interest of this approach is that it can be used for analyzing the consequences of parameter uncertainty on model predictions.

In the Bayesian approach, the unknown model parameters are treated as random variables distributed according to a probability distribution, which expresses uncertainty about parameter values. Prior to collecting data, this distribution is based on the existing knowledge about the parameter values before the measurement of new data. If no information is available, it is possible to define an uninformative prior distribution. However, many of the parameters of agronomic models are directly meaningful and it is often possible to specify lower and upper bounds for these parameters. In such cases, the prior parameter distribution can be defined, for instance, by a uniform distribution. Posterior to collecting data, the parameter distribution is given by Bayes' theorem. The posterior parameter distribution depends both on the prior distribution and on the measurements. Thus, this distribution contains all the available information about the model parameters. In practice, we would then like to summarize the posterior parameter distribution.

Due to the complexity of agronomic models (non-linearity, high number of parameters), it is not possible to calculate analytically the posterior parameter distribution. However, the growing power of computers and the development of new methods make the Bayesian approach accessible even with complex models [4, 6, 12, 15]. One of these methods is the Metropolis-Hastings algorithm [10, 11, 16]. This is a technique for generating samples from high-dimensional distributions, here the posterior parameter distribution. The principle is to generate a large enough sample from the posterior parameter distribution so that features of this distribution (expected parameter values, parameter variances...) can be accurately determined. Another increasingly popular method is the Generalized Likelihood Uncertainty Estimation method (GLUE) [3, 6, 18]. The principle of this method is to discretize the parameter space by generating a large number of parameter values from the prior distribution. Weights are then calculated at each parameter value from likelihood and prior density values. The Metropolis-Hastings algorithm and the GLUE method have already been applied to complex models [4, 6, 12, 18]. However, no study has been performed yet for comparing the performances of these methods. The purpose of this paper is to compare the performances of the Metropolis-Hastings algorithm and the GLUE method for estimating the parameters of a complex model. The considered model has the main features of an agronomic model; it is a non-linear model that includes a high number of parameters (22). The model is inspired by the model of Makowski et al. [13, 14] that predicts the response of yield to applied nitrogen. The Metropolis-Hastings algorithm and the GLUE method are used to obtain parameter estimates, which are then evaluated and compared. We also present the interest of the Bayesian approach for studying the consequences of parameter uncertainty on predictions of agronomic models.

## 2. MATERIALS AND METHODS

### 2.1. The Bayesian approach

The posterior parameter distribution is given by Bayes' theorem:

$$\pi(\theta|Y) = \pi(Y|\theta)\ \pi(\theta) / \pi(Y)$$

where $Y$ is the vector of measurements, $\pi(\theta)$ is the prior parameter distribution, $\pi(\theta|Y)$ is the posterior parameter distribution, $\pi(Y)$ is a constant of proportionality determined by the requirement that the integral of $\pi(\theta|Y)$ over the parameter space equals 1, and $\pi(Y|\theta)$ is a likelihood function. The likelihood is the probability of the data $Y$ given the parameters $\theta$. Its value is determined from the probability distribution of the errors between modeled and observed data. It is readily seen that both the prior distribution and the new data affect the posterior parameter distribution.

### 2.2. The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm [10, 11, 16] is a Markov chain Monte Carlo technique for generating

samples from the posterior distribution $\pi(\theta|Y)$. The algorithm starts with a vector value $\theta_0$. Then a sequence of $N$ parameter vectors $\theta_i$, $i$=1, ..., $N$, is generated as follows:

(1) Generate a candidate parameter vector $\theta^*$ from a proposal distribution $q(\theta^*|\theta_{i-1})$, for instance a Normal distribution with mean equal to $\theta_{i-1}$.

(2) Calculate $T = \dfrac{\pi(Y|\theta^*)\pi(\theta^*)q(\theta_{i-1}|\theta^*)}{\pi(Y|\theta_{i-1})\pi(\theta_{i-1})q(\theta^*|\theta_{i-1})}$, where $\pi(Y|\theta^*)$ and $\pi(Y|\theta_{i-1})$ are the likelihood values of the parameter vectors $\theta^*$ and $\theta_{i-1}$ respectively, and $\pi(\theta^*)$ and $\pi(\theta_{i-1})$ are the prior densities of $\theta^*$ and $\theta_{i-1}$ respectively.

(3) If $min(1, T) > u$, where $u$ is drawn from a uniform distribution on the interval (0, 1), then $\theta_i = \theta^*$ else $\theta_i = \theta_{i-1}$.

After an initial phase of say $M$ iterations, the chain thus constructed will converge to a chain with elements randomly drawn from the posterior parameter distribution $\pi(\theta|Y)$. The $M$ first iterations should be discarded.

Before using the Metropolis-Hastings algorithm, it is necessary to choose the starting value $\theta_0$, the proposal distribution $q(\theta^*|\theta_{i-1})$, the total number of iterations $N$, and the number of discarded iterations $M$. The definition of precise rules for choosing these elements is still an active area of research. According to Gilks et al. [10], the choice of $\theta_0$ is not very critical. On the other hand, the choice of the proposal distribution $q(\theta^*|\theta_{i-1})$ is an important issue. A common practice is to use a Normal distribution with mean $\theta_{i-1}$ and constant covariance matrix $\Sigma$. This is equivalent to assuming that $\theta^*|\theta_{i-1} \sim N(\theta_{i-1}, \Sigma)$. Several authors [4, 12] suggest choosing $\Sigma$ such that the acceptance rate of the test performed in step (2) of the algorithm is in the range 20–70%. This rule will be used in the application presented in this paper. The choice of the numbers $M$ (number of iterations to be discarded) and $N$ (total number of iterations) is not straightforward. In this paper, we use $M$=$N$/2, which is much higher than the 1 or 2% suggested by Geyer [9] and than the 20% suggested by Raftery et al. [17]. Thus, we avoid keeping parameter values which might be far from the posterior distribution. The most obvious method for determining $N$ is to run several chains with different starting values and different lengths and then to compare the results obtained with these different chains, for instance the average value of the generated parameters [8]. If the results do not agree adequately, the number of iterations $N$ must be increased. Such sensitivity analysis will be performed in our application.

The generated sample can be used for summarizing features of the posterior parameter distribution. For instance, the sample of parameter vectors $\theta_i$, $i$=$M$+1, ..., $N$, can be used to calculate the posterior means, which minimizes expected quadratic loss [2]:

$$\bar{\theta} = \frac{1}{N-M}\sum_{i=M+1}^{N}\theta_i \cdot$$

The vector $\bar{\theta}$ can then be considered as an estimate of the model parameters. Calculating parameter estimates involves a loss of information compared to the full posterior probability distribution. The sample of parameter vectors can also be used for calculating posterior variances, correlations between parameters, and distribution of model prediction.

### 2.3. Generalized likelihood uncertainty estimation (GLUE)

The principle of the GLUE method [3, 6, 18] is to approximate the posterior parameter distribution $\pi(\theta|Y)$ by a discrete probability distribution $(\theta_i, p_i)$, $i$=1, ..., $N$, $\sum_{i=1}^{N} p_i = 1$, where $p_i$ is the probability associated with the parameter vector $\theta_i$. The method proceeds as follows:

(1) Randomly generate $N$ vectors $\theta_i$, $i$=1, ..., $N$, from the prior parameter distribution $\pi(\theta)$.

(2) Calculate the likelihood values $\pi(Y|\theta_i)$ and the prior density $\pi(\theta_i)$, $i$=1, ..., $N$, associated with the different generated parameter vectors.

(3) Calculate $p_i = \dfrac{\pi(Y|\theta_i)\pi(\theta_i)}{\sum_{j=1}^{N}\pi(Y|\theta_j)\pi(\theta_j)}$.

The pairs $(\theta_i, p_i)$, $i$=1, ..., $N$, can be used to determine various characteristics of the posterior distribution, for instance, the posterior means $\bar{\theta} = \sum_{i=1}^{N} p_i\theta_i$. The application of the GLUE procedure simply requires the definition of the total number of generated parameter vectors $N$ and so is more straightforward than the application of the Metropolis-Hastings algorithm. However, no study has yet been performed to compare the performances of the Metropolis-Hastings algorithm and the GLUE method. Such a study is presented below in the case of a model including a high number of parameters.

### 2.4. The model

We consider a linear-plus-plateau model that relates observation $y$ to explanatory variable $x$ as follows:

$$y = Y_{MAX} + Q\,(x - X_{MAX}) + \varepsilon \quad \text{if } x < X_{MAX} \qquad (1)$$

$$y = Y_{MAX} + \varepsilon \quad \text{if } x \geq X_{MAX} \qquad (2)$$

where $Y_{MAX}$ is the maximal value of $y$, $X_{MAX}$ is the $x$-value at which $Y_{MAX}$ is reached, $Q$ is the slope of the linear part of the response, and $\varepsilon$ is the error term. This type of model was used by Makowski et al. [13, 14] to predict the response of yield to applied nitrogen. The authors showed that the parameter values of the linear-plus-plateau function vary between site-years and suggested that the values of $Y_{MAX}$, $X_{MAX}$ and $Q$ depend linearly on site-year characteristics. In order to increase the number of parameters of the model, we assume that $Y_{MAX}$, $X_{MAX}$ and $Q$ depend on site-year characteristics as follows:

$$Y_{MAX} = a_0 + \sum_{s=1}^{6} a_s T_s \qquad (3)$$

$$Q = b_0 + \sum_{s=1}^{6} b_s U_s \qquad (4)$$

$$X_{MAX} = c_0 + \sum_{s=1}^{6} c_s V_s \qquad (5)$$

where $T_s$, $U_s$, $V_s$, $s$=1, ..., 6 represent some site-year characteristics and $a_s$, $b_s$, $c_s$, $s$ = 0, ..., 6 are parameters. Each variable is assumed to be distributed as N(0, 1) and to be independent of the others, except $T_1$ and $T_2$ which are supposed to be related to each other by $T_1 = T_2 + \eta$ where $\eta \sim N(0, \tau^2)$ and $\tau^2 = 0.0203$. As a consequence, $T_1$ and $T_2$ are strongly correlated, $\mathrm{cor}(T_1, T_2) = 0.99$. Finally, we assume independent and normally distributed error terms, $\varepsilon \sim N(0, \sigma^2)$, where $\sigma$ is the error term standard deviation.

Let $\theta$ denote the vector of the model parameters. $\theta$ includes 22 elements, $\theta = [a_0, a_1, a_2, a_3, a_4, a_5, a_6, b_0, b_1, b_2, b_3, b_4, b_5, b_6, c_0, c_1, c_2, c_3, c_4, c_5, c_6, \sigma]^t$, where $^t$ denotes the transpose matrix operator. The true values of the elements of $\theta$ are shown in Table I. Note that the true values of $a_6$, $b_6$, and $c_6$ are equal to zero and so $T_6$, $U_6$, and $V_6$ are not related to $y$.

The model (1–5), inspired by Makowski et al. [13, 14], does not represent a real model, but has certain important characteristics in common with crop models. In particular the model is non-linear in its parameters, there are many parameters, and two of the variables ($T_1$ and $T_2$) are highly correlated with each other.

### 2.5. Prior parameter distributions

Two prior distributions are defined. Both distributions assume the parameters to be a priori independent and uniformly distributed. The means and variances of the first prior distribution are given in Table I. We see that the prior means are different from the true parameter values, as would usually be the case with an agronomic model. The difference is important for some parameters (for instance the relative difference between the prior mean and the true value of $a_0$ is equal to 0.25) but marginal for others. The lower and upper bounds of the first prior distribution are fixed respectively to 50% and 150% of the prior means for all parameters except $a_6$, $b_6$, and $c_6$ and respectively to 0 and 200% of the prior means for $a_6$, $b_6$, and $c_6$. With these bounds, all the true parameter values are in the ranges of parameter values covered by the prior distribution. We will refer to this prior distribution as the initial prior distribution.

A second prior distribution is defined in order to study the consequences of wrong assumptions about what might be the parameter values. This second prior distribution is based on the same prior means as those given in Table I but differs from the initial prior distribution by its upper and lower bounds which are fixed here respectively to 75% and 125% of the prior means for all parameters except $a_6$, $b_6$, and $c_6$ and respectively to 50% and 150% of the prior means for $a_6$, $b_6$, and $c_6$. This second prior distribution has lower variances than the initial prior distribution and does not include the true values of six parameters, namely $a_2$, $a_6$, $b_6$, $c_2$, $c_4$, and $c_6$. We will refer to this second prior distribution as the prior distribution with reduced variances.

### 2.6. Generation of data and calculation of posterior parameter distributions

One hundred samples of size 5 and 40 $y$ data are generated by using the model defined in Section 2.4 with parameters fixed to their true values and by randomly drawing values of $\varepsilon$, $x$, $T_s$, $U_s$, and $V_s$, $s$=1, ..., 6; $x$ values are drawn from a uniform distribution in the range 0–400. Posterior parameter distributions are derived for all generated samples of data by using the Metropolis-Hastings algorithm and the GLUE method. As the model errors $\varepsilon$ are assumed independent and normally distributed, we use the following likelihood function:

$$\pi(Y|\theta) = \prod_{j=1}^{K} (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{[y_j - f(\omega_j; \theta)]^2}{2\sigma^2} \right\}$$

**Table I.** Mean (M) and standard deviation (SD) of the prior parameter distribution compared to the true parameter values (T).

| | Prior distribution | | True parameter value | Difference |
|---|---|---|---|---|
| | Mean (M) | SD | (T) | \|M–T\|/T |
| $a_0$ | 12.01 | 3.468 | 9.61 | 0.25 |
| $a_1$ | 0.64 | 0.185 | 0.51 | 0.26 |
| $a_2$ | 0.37 | 0.108 | 0.51 | 0.27 |
| $a_3$ | 0.30 | 0.087 | 0.25 | 0.19 |
| $a_4$ | 0.10 | 0.030 | 0.10 | 0.03 |
| $a_5$ | 0.05 | 0.013 | 0.05 | 0.08 |
| $a_6$ | –0.09 | 0.050 | 0 | 0.09[a] |
| $b_0$ | 0.029 | 0.0083 | 0.032 | 0.11 |
| $b_1$ | 0.00116 | 0.00034 | 0.00093 | 0.25 |
| $b_2$ | 0.00120 | 0.00035 | 0.00093 | 0.28 |
| $b_3$ | 0.00042 | 0.00012 | 0.00047 | 0.09 |
| $b_4$ | 0.00023 | 0.00007 | 0.00019 | 0.24 |
| $b_5$ | 0.00008 | 0.00002 | 0.00009 | 0.14 |
| $b_6$ | 0.070 | 0.041 | 0 | 0.07[a] |
| $c_0$ | 129.40 | 37.35 | 133.00 | 0.03 |
| $c_1$ | 20.47 | 5.91 | 21.88 | 0.06 |
| $c_2$ | 16.62 | 4.80 | 21.88 | 0.24 |
| $c_3$ | 11.21 | 3.24 | 10.94 | 0.02 |
| $c_4$ | 3.32 | 0.96 | 4.38 | 0.24 |
| $c_5$ | 2.71 | 0.78 | 2.19 | 0.24 |
| $c_6$ | –0.055 | 0.032 | 0 | 0.06[a] |
| $\sigma$ | 0.3 | 0.09 | 0.4 | 0.25 |

[a] The difference between M and T is measured by |M–T| for these parameters.

where $y_j$ is the $j$th $y$ value in the data set $Y$, $\omega_j$ is the vector of explanatory variables (i.e. the vector including the value of $x$ and the values of the 18 site-year characteristics) associated with $y_j$, $f(\omega_j;\theta)$ is the model prediction of $y_j$ defined by (1) to (5), and $K$ is the total number of data in the data set (5 or 40). Note that other expressions of the likelihood can be used for implementing the GLUE method [18]. The Metropolis-Hastings algorithm is implemented by using two starting vectors $\theta_0$ successively. The elements of the first starting vector are the means of the prior distributions. In the second starting vector, the values of $a_0$, $b_0$, and $c_0$ are deviated respectively by –6, +0.01, and +20 from their prior means. The proposal distribution used for generating candidate parameter values in step (1) of the Metropolis-Hastings algorithm is a Normal distribution with mean $\theta_{i-1}$, constant variances, and zero covariances. The variances of the proposal distribution are chosen so the acceptance rate of the test performed in step (2) of the algorithm is in the range 20–70%. Several chains are run successively with different proposal distribution variances until the acceptance rate is satisfactory. The number of iterations $N$ of the Metropolis-Hastings algorithm and of the GLUE method is first fixed to 100 000 and then to 200 000. Only the last

50% of the parameter vectors generated with the Metropolis-Hastings algorithm are retained. All the calculations are done with Fortran 90 [5].

### 2.7. Evaluation of parameter estimates and of model predictions

Each posterior distribution is used to calculate the vector of the posterior means $\bar{\theta}$. Then, $\bar{\theta}$ is considered as an estimate of the model parameter vector $\theta$. The quality of $\bar{\theta}$ is evaluated by calculating, for each parameter except $a_6$, $b_6$, and $c_6$, the relative absolute difference between the parameter estimate and the true parameter value, that is the ratio of the absolute difference to the true parameter value. For evaluating the estimated values of $a_6$, $b_6$, and $c_6$, we use simply the absolute differences between the parameter estimates and the true values because the true values are zero for these three parameters. As parameter estimators of agronomic models are often used for predicting $y$, it is interesting to evaluate the accuracy of model predictions when the parameters are fixed to their estimated values. Following Harmon and Challenor [12], predicted values of $y$ are obtained with our model by using $\bar{\theta}$ as parameter estimators. The accuracy of the model predictions is then evaluated by calculating values of mean squared error of prediction (*MSEP*). The mean squared error of prediction associated with each vector $\bar{\theta}$ is estimated from 10 000 data $y$ generated by using the model with the true parameter values:

$$\hat{MSEP}(\bar{\theta}) = \frac{1}{10000} \sum_{j=1}^{10\,000} [y_j - f(\omega_j; \bar{\theta})]^2$$

where $f(\omega_j;\bar{\theta})$ is the model prediction when model parameters are fixed to their posterior means $\bar{\theta}$. Gelfand et al. [7] have proposed another approach to evaluate accuracy of model predictions. They propose estimating the value of mean squared error of prediction $MSEP(\theta)$ for each value of $\theta$ taken in the posterior parameter distribution. A posterior distribution for *MSEP* can then be obtained. This method may be more informative but is not applied here.

## 3. RESULTS

### 3.1. Posterior parameter distributions

Tables II to V give estimated values of the means and standard deviations of several posterior parameter distributions. These results were obtained by using the Metropolis-Hastings algorithm and the GLUE method with a sample of five observations and with a sample of 40 observations. In all cases, the initial prior distribution described in Section 2.5 was used. The results were found to be quite insensitive to the starting values and to the total number of iterations. The results presented in Tables II to V were obtained by using the prior means as starting values and by generating 100 000 parameter vectors. The first 50 000 parameter vectors generated with the Metropolis-Hastings algorithm were discarded.

Table II gives the estimated means and standard deviations of the posterior parameter distribution obtained by using the Metropolis-Hastings algorithm with five data samples. The differences between the estimated posterior means and the true parameter values allow us to evaluate the improvement in parameter estimation resulting from using five observations. Thus, Table II shows that the posterior means of $a_0$ and $b_6$ are quite different from the prior means and are nearer to the true values. The relative differences between the posterior means of $a_0$ and $b_6$ and the true parameter values are equal only to 0.003 and 0.005 respectively (Tab. II) whereas the relative differences between the prior means and the true values are equal to 0.25 and 0.07 respectively (Tab. I). Moreover, the standard deviations of the posterior distributions of $a_0$ and $b_6$ are only equal to 0.26 and 0.0033 respectively (Tab. II), and so are much smaller than the standard deviations of the prior distributions (Tab. I). These results show that the use of five data samples has improved the estimation of the parameters $a_0$ and $b_6$ and has reduced the uncertainty about the values of these two parameters. On the contrary, for the other parameters, the relative differences between the posterior means and the true values are not much smaller and, sometimes, are even higher than the relative differences between the prior means and the true values. Moreover, the standard deviations of the posterior distributions of these parameters (and so the level of uncertainty) remain high.

Table III gives the estimated means and standard deviations of a posterior parameter distribution obtained by using the Metropolis-Hastings algorithm with 40 data samples. Compared to the results obtained with 5 data samples (Tab. II), the use of additional data samples improves the estimation of several parameters, notably of $b_6$, $c_0$, and $\sigma$. Thus, the posterior means of $b_6$, $c_0$, and $\sigma$ are equal to 0.00235, 131.48, and 0.42 respectively when 40 data samples are used (Tab. III). These values are not very different from the true values of $b_6$, $c_0$, and $\sigma$ respectively equal to 0, 133, and 0.4 (Tab. I). For several parameters, the variances are smaller when 40 data samples are used

**Table II.** Mean (M) and standard deviation (SD) of the posterior parameter distribution obtained with five data samples by using the Metropolis-Hastings algorithm.

| | Posterior distribution | | True parameter value | Difference |
|---|---|---|---|---|
| | Mean (M) | SD | (T) | \|M–T\|/T |
| $a_0$ | 9.64 | 0.26 | 9.61 | 0.003 |
| $a_1$ | 0.66 | 0.15 | 0.51 | 0.31 |
| $a_2$ | 0.42 | 0.091 | 0.51 | 0.17 |
| $a_3$ | 0.31 | 0.092 | 0.25 | 0.21 |
| $a_4$ | 0.085 | 0.025 | 0.10 | 0.16 |
| $a_5$ | 0.048 | 0.013 | 0.05 | 0.06 |
| $a_6$ | –0.051 | 0.033 | 0 | 0.05[a] |
| $b_0$ | 0.034 | 0.00555 | 0.032 | 0.07 |
| $b_1$ | 0.00113 | 0.00032 | 0.00093 | 0.21 |
| $b_2$ | 0.00179 | 0.00036 | 0.00093 | 0.92 |
| $b_3$ | 0.00042 | 0.00012 | 0.00047 | 0.09 |
| $b_4$ | 0.00025 | 0.00007 | 0.00019 | 0.33 |
| $b_5$ | 0.00010 | 0.00003 | 0.000093 | 0.05 |
| $b_6$ | 0.00536 | 0.00326 | 0 | 0.005[a] |
| $c_0$ | 166.17 | 18.26 | 133.00 | 0.25 |
| $c_1$ | 20.30 | 5.98 | 21.88 | 0.07 |
| $c_2$ | 18.11 | 4.77 | 21.88 | 0.17 |
| $c_3$ | 10.51 | 2.77 | 10.94 | 0.04 |
| $c_4$ | 3.21 | 0.95 | 4.38 | 0.27 |
| $c_5$ | 2.61 | 0.76 | 2.19 | 0.19 |
| $c_6$ | –0.063 | 0.03 | 0 | 0.06[a] |
| $\sigma$ | 0.3 | 0.09 | 0.4 | 0.26 |

[a] The difference between M and T is measured by |M–T| for these parameters.

than when 5 data samples are used. For instance, the standard deviation of $c_0$ is equal to 18.26 with 5 data samples (Tab. II) but is reduced to 6.16 with 40 data samples (Tab. III). Thus, the use of 35 additional observations decreases the levels of uncertainty about the values of some parameters.

Tables IV and V give the means and standard deviations of two posterior parameter distributions obtained with the GLUE method. The results are quite similar to the results obtained with the Metropolis-Hastings algorithm. Thus, with the GLUE method also, the use of data strongly improves the estimation of the parameters $a_0$ and $b_6$. The differences between the posterior means of $a_0$ and $b_6$ obtained by using GLUE and the true parameter values (Tabs. IV and V) are much smaller than the differences between the prior means and the true parameter values (Tab. I). The standard deviations of the posterior distributions of $a_0$ and $b_6$ obtained with the GLUE method are also smaller (Tabs. IV and V) than the standard deviations of the prior parameter distribution (Tab. I). Like the Metropolis-Hastings algorithm, the

**Table III.** Mean (M) and standard deviation (SD) of the posterior parameter distribution obtained with forty data samples by using the Metropolis-Hastings algorithm.

| | Posterior distribution | | True parameter value | Difference |
|---|---|---|---|---|
| | Mean (M) | SD | (T) | \|M–T\|/T |
| $a_0$ | 9.67 | 0.09 | 9.61 | 0.01 |
| $a_1$ | 0.60 | 0.12 | 0.51 | 0.18 |
| $a_2$ | 0.38 | 0.11 | 0.51 | 0.26 |
| $a_3$ | 0.26 | 0.05 | 0.25 | 0.01 |
| $a_4$ | 0.11 | 0.03 | 0.10 | 0.05 |
| $a_5$ | 0.040 | 0.010 | 0.05 | 0.21 |
| $a_6$ | –0.044 | 0.030 | 0 | 0.04[a] |
| $b_0$ | 0.037 | 0.0029 | 0.032 | 0.15 |
| $b_1$ | 0.00108 | 0.00027 | 0.00093 | 0.16 |
| $b_2$ | 0.00121 | 0.00037 | 0.00093 | 0.30 |
| $b_3$ | 0.00042 | 0.00012 | 0.00047 | 0.09 |
| $b_4$ | 0.00021 | 0.000073 | 0.00019 | 0.12 |
| $b_5$ | 0.00010 | 0.000018 | 0.000093 | 0.02 |
| $b_6$ | 0.00235 | 0.00153 | 0 | 0.002[a] |
| $c_0$ | 131.48 | 6.16 | 133.00 | 0.01 |
| $c_1$ | 23.14 | 3.13 | 21.88 | 0.06 |
| $c_2$ | 20.73 | 2.03 | 21.88 | 0.05 |
| $c_3$ | 9.75 | 2.45 | 10.94 | 0.11 |
| $c_4$ | 3.27 | 0.98 | 4.38 | 0.25 |
| $c_5$ | 2.14 | 0.48 | 2.19 | 0.02 |
| $c_6$ | –0.054 | 0.035 | 0 | 0.05[a] |
| $\sigma$ | 0.42 | 0.02 | 0.4 | 0.05 |

[a] The difference between M and T is measured by |M–T| for these parameters.

GLUE method does not improve the estimation of the other parameters much.

### 3.2. Mean squared error of prediction

Table VI shows *MSEP* values obtained when the model parameters are fixed to the means of the posterior parameter distribution. The posterior means were derived by using the Metropolis-Hastings algorithm and the GLUE method with 100 000 or 200 000 generated parameter vectors, with two different starting values, and with 5 or 40 data samples. The results reported in Table VI were not obtained for one particular sample of data samples *Y*, but for 100 samples of data. For each sample, the vector of the means of the posterior parameter distribution, $\bar{\theta}$, was derived either by using the Metropolis-Hastings algorithm or the GLUE method. Then, the *MSEP* value associated with each $\bar{\theta}$ value was calculated as explained in Section 2.7. Finally, the expected values of the *MSEP* across the 100 samples of data were estimated from the individual *MSEP* values.

**Table IV.** Mean (M) and standard deviation (SD) of the posterior parameter distribution obtained with five data samples by using GLUE.

| | Posterior distribution | | True parameter value | Difference |
|---|---|---|---|---|
| | Mean (M) | SD | (T) | \|M–T\|/T |
| $a_0$ | 9.63 | 0.31 | 9.61 | 0.002 |
| $a_1$ | 0.70 | 0.11 | 0.51 | 0.38 |
| $a_2$ | 0.39 | 0.100 | 0.51 | 0.23 |
| $a_3$ | 0.32 | 0.079 | 0.25 | 0.24 |
| $a_4$ | 0.101 | 0.026 | 0.10 | 0.01 |
| $a_5$ | 0.049 | 0.013 | 0.05 | 0.05 |
| $a_6$ | –0.088 | 0.052 | 0 | 0.09[a] |
| $b_0$ | 0.037 | 0.00391 | 0.032 | 0.15 |
| $b_1$ | 0.00122 | 0.00031 | 0.00093 | 0.31 |
| $b_2$ | 0.00130 | 0.00033 | 0.00093 | 0.39 |
| $b_3$ | 0.00047 | 0.00010 | 0.00047 | 0.01 |
| $b_4$ | 0.00022 | 0.00007 | 0.00019 | 0.18 |
| $b_5$ | 0.00009 | 0.00002 | 0.000093 | 0.07 |
| $b_6$ | 0.00640 | 0.00315 | 0 | 0.01[a] |
| $c_0$ | 163.67 | 17.99 | 133.00 | 0.23 |
| $c_1$ | 21.44 | 6.04 | 21.88 | 0.02 |
| $c_2$ | 16.65 | 4.55 | 21.88 | 0.24 |
| $c_3$ | 11.14 | 3.18 | 10.94 | 0.02 |
| $c_4$ | 3.42 | 0.90 | 4.38 | 0.22 |
| $c_5$ | 2.77 | 0.69 | 2.19 | 0.27 |
| $c_6$ | –0.049 | 0.03 | 0 | 0.05[a] |
| $\sigma$ | 0.3 | 0.07 | 0.4 | 0.19 |

[a] The difference between M and T is measured by |M–T| for these parameters.

The first line of Table VI shows that, when the parameters are fixed to their prior means (Tab. I), the *MSEP* of the model is equal to 17.17. This value is much higher than the residual error variance $\sigma^2$, equal to 0.16, which represents the smallest possible *MSEP* of the model, that is the *MSEP* of the model with the true parameter values.

The *MSEP* is strongly decreased when the model parameters are fixed to the means of posterior parameter distribution calculated by using the Metropolis-Hastings algorithm or the GLUE method. When the initial prior parameter distribution defined in Section 2.5 is used, the expected *MSEP* value is between 2.81 and 3.82 with five data samples, and is between 0.21 and 0.27 with forty data samples (Tab. VI).

The smallest *MSEP* value (0.21) is obtained with the Metropolis-Hastings algorithm and 200 000 iterations. This value is very near from the smallest possible *MSEP* value, 0.16. The starting values have no significant effect on the *MSEP*. The number of iterations has a small effect on the *MSEP* value when five data samples are used, but no effect when forty data samples are used. The *MSEP* value is not strongly influenced by the Bayesian method

**Table V.** Mean (M) and standard deviation (SD) of the posterior parameter distribution obtained with forty data samples by using GLUE.

| | Posterior distribution | | True parameter value | Difference |
|---|---|---|---|---|
| | Mean (M) | SD | (T) | \|M–T\|/T |
| $a_0$ | 9.91 | 0.05 | 9.61 | 0.03 |
| $a_1$ | 0.43 | 0.02 | 0.51 | 0.15 |
| $a_2$ | 0.44 | 0.03 | 0.51 | 0.13 |
| $a_3$ | 0.20 | 0.01 | 0.25 | 0.20 |
| $a_4$ | 0.11 | 0.01 | 0.10 | 0.08 |
| $a_5$ | 0.058 | 0.003 | 0.05 | 0.14 |
| $a_6$ | –0.033 | 0.049 | 0 | 0.03[a] |
| $b_0$ | 0.038 | 0.0055 | 0.032 | 0.19 |
| $b_1$ | 0.00118 | 0.00023 | 0.00093 | 0.26 |
| $b_2$ | 0.00130 | 0.00017 | 0.00093 | 0.39 |
| $b_3$ | 0.00036 | 0.00004 | 0.00047 | 0.24 |
| $b_4$ | 0.00022 | 0.000040 | 0.00019 | 0.16 |
| $b_5$ | 0.00011 | 0.000013 | 0.000093 | 0.22 |
| $b_6$ | 0.00308 | 0.00063 | 0 | 0.003[a] |
| $c_0$ | 136.53 | 10.36 | 133.00 | 0.03 |
| $c_1$ | 22.53 | 0.40 | 21.88 | 0.03 |
| $c_2$ | 23.81 | 1.33 | 21.88 | 0.09 |
| $c_3$ | 10.61 | 0.75 | 10.94 | 0.03 |
| $c_4$ | 2.40 | 0.36 | 4.38 | 0.45 |
| $c_5$ | 3.45 | 0.81 | 2.19 | 0.58 |
| $c_6$ | –0.052 | 0.010 | 0 | 0.05[a] |
| $\sigma$ | 0.36 | 0.04 | 0.4 | 0.10 |

[a] The difference between M and T is measured by |M–T| for these parameters.

used to derive the posterior parameter distribution. However, the expected *MSEP* values obtained with the GLUE method are slightly higher when the number of data samples is 40 (Tab. VI). On the contrary, the prior parameter distribution has a strong influence on *MSEP* values. Thus, when the prior distribution with reduced variances is used, the expected *MSEP* values obtained with 5 and 40 data samples are equal to 5.67 and 1.86 respectively. These *MSEP* are much higher than the values obtained with the initial prior distribution (Tab. VI). This is due to the fact that parameter space covered by the prior distribution with reduced variances does not include the true values of six parameters (see Sect. 2.5). Consequently, the posterior means of these parameters remain relatively different from the true parameter values, even if a high number of data samples is used.

### 3.3. Analysis of model prediction uncertainty

The results presented above show the interest of the Bayesian approach for estimating model parameters.

**Table VI**. Expected value of the mean squared error of predictions ($\hat{E}[\hat{MSEP}(\bar{\theta})]$) obtained by using different Bayesian methods, different prior distributions, and different amounts of data. Calculations were performed by using 100 samples of data.

| Nb. of data | Algorithm | Nb. of iterations | Starting values | $\hat{E}[\hat{MSEP}(\bar{\theta})]$ |
|---|---|---|---|---|
| | | *Initial prior parameter distribution* | | |
| 0 | – | – | – | 17.71 |
| 5 | Metropolis | 100000 | 1[a] | 3.82 (0.4)[c] |
| 5 | Metropolis | 100000 | 2[b] | 3.74 (0.4) |
| 5 | Glue | 100000 | – | 3.75 (0.42) |
| 5 | Metropolis | 200000 | 1 | 2.81 (0.34) |
| 5 | Metropolis | 200000 | 2 | 2.97 (0.37) |
| 5 | Glue | 200000 | – | 3.19 (0.38) |
| 40 | Metropolis | 100000 | 1 | 0.22 (0.005) |
| 40 | Metropolis | 100000 | 2 | 0.22 (0.003) |
| 40 | Glue | 100000 | – | 0.27 (0.005) |
| 40 | Metropolis | 200000 | 1 | 0.22 (0.003) |
| 40 | Metropolis | 200000 | 2 | 0.21 (0.001) |
| 40 | Glue | 200000 | – | 0.26 (0.005) |
| | | *Prior parameter distribution with reduced variances* | | |
| 5 | Metropolis | 100000 | 1 | 5.67 (0.36) |
| 40 | Metropolis | 100000 | 1 | 1.86 (0.02) |

[a] Number 1 indicates that the starting values are the prior means for all parameters.
[b] Number 2 indicates that, for some parameters, the starting values are different from the prior means (see text).
[c] Standard error of the estimator of $E[MSEP(\theta)]$ is between parentheses.

Another interest of the Bayesian approach is that it can be used for analyzing the prediction uncertainty due to the uncertainty about parameter values. This point is illustrated in Figures 1 and 2. These figures show probability distributions of model predictions. Each distribution is based on 50 000 parameter vectors drawn either from a prior parameter distribution or from a posterior parameter distribution. In the latter case, the parameter values are generated by using the Metropolis-Hastings algorithm with 5 or 40 observations. Each generated parameter vector is used in turn for calculating the model prediction of $y$ when all the explanatory variables are 0. Thus, each one of the reported distributions can be expressed by $\{f(0;\theta_k), k=1, ..., 50\ 000\}$, where $\theta_k$ is the $k$th parameter vector drawn either from a prior parameter distribution or from a posterior parameter distribution.

Distributions reported in Figure 1 are obtained by using the initial prior parameter distribution. The thick line represents the distribution of model predictions when the

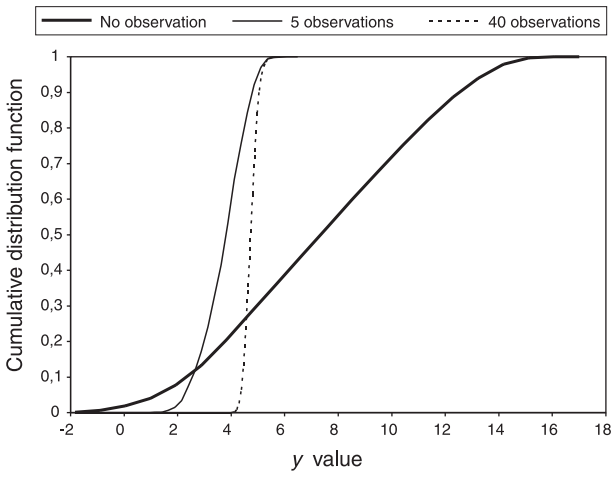


**Figure 1.** Cumulative distribution functions of $y$ value when all the explanatory variables are zero. Functions were obtained by using the initial prior parameter distribution and by using 0, 5, or 40 observations.

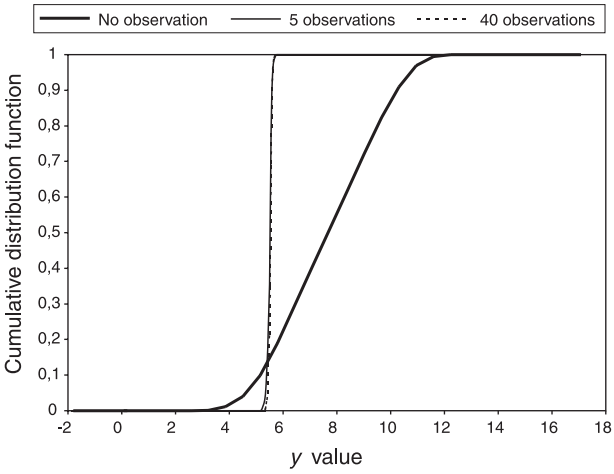


**Figure 2.** Cumulative distribution functions of *y* value when all the explanatory variables are zero. Functions were obtained by using the prior parameter distribution with reduced variances and by using 0, 5, or 40 observations.

parameter vectors are directly drawn from the initial prior parameter distribution. The other two distributions are generated by using posterior parameter distributions obtained with 5 and 40 observations. The prior parameter distribution used to derive the two posterior distributions was the initial prior distribution. When 5 or 40 observations are used, the standard deviations of the distributions of model predictions are equal to 0.83 and 0.24 respectively. When no observations are used, the standard deviation is equal to 3.8 and so is much higher. These results show that the uncertainty about model predictions is strongly decreased when model predictions are derived from a posterior parameter distributions rather than from a prior parameter distribution, even if only five data samples are used to calculate the posterior parameter distribution.

Figure 2 shows the effect of a reduction of the prior variances on the level of uncertainty about model predictions. The distributions reported in this figure are obtained by using the prior parameter distribution with reduced variances (see Sect. 2.5). In Figure 2, the standard deviations of model predictions are equal respectively to 1.9, 0.08 and 0.06 when 0, 5 and 40 are used. These values are much lower than the standard deviations of the model predictions based on the initial prior parameter distribution (Fig. 1). This result shows that the prior parameter distribution has a strong influence on the level of uncertainty about model predictions. If one defines a prior parameter distribution with small variances (i.e. a very informative prior), the level of uncertainty about model predictions is likely to be low. The definition of an unrealistically informative prior can thus lead the user to put too much confidence in the model predictions.

## 4. DISCUSSION

This paper has shown that the Metropolis-Hastings algorithm and the GLUE procedure are two interesting methods for estimating the parameters of agronomic models. A first interest of these methods is that they give satisfactory parameter estimates, even with little data (which is the usual situation with agronomic models). This is due to the fact that the parameter estimates calculated by using Bayesian methods are based on two types of information: expert knowledge about parameter values and data from experimentation. Expert knowledge is taken into account through a prior parameter distribution. This prior distribution prevents the model parameters from taking unrealistic values. A second interest of Bayesian methods is that they are useful for analyzing the uncertainty about parameter values. In the Bayesian approach, the parameters are assumed to be random variables. The probability distributions of these variables indicate the extent of the uncertainty about model parameter values. Such representation of the parameter uncertainty can be useful for taking into account risk in model-based decision-making.

The two methods presented in this paper have been found to be equally good in terms of *MSEP* value. However, we advise using the Metropolis-Hastings algorithm rather than the GLUE method. Indeed, mathematical properties ensure that, once the algorithm has settled down, the sequence of parameters generated with the Metropolis-Hastings algorithm will have the same distribution as the posterior parameter distribution even if the number of parameters is high [4, 10]. The GLUE method proceeds by a discretization of the parameter space. This approach can lead to an inaccurate representation of the posterior parameter distribution when the model parameters are numerous.

In the Bayesian approach, the data are taken into account through a likelihood function. This function plays an important role in the two methods presented in this paper. The likelihood function is determined by the hypothesis made on the model errors. In our simulation study, we have assumed that the model errors were all

independent. This hypothesis may be unrealistic with some data sets. For instance, fertilizer trials generally constitute several measurements taken on different site-years and at different fertilizer doses. The measurements taken on the same site-year cannot be reasonably considered as independent [19]. The likelihood function used in our application is not suitable in such situations and should be modified. Bennett et al. [1] explain how to apply the Bayesian approach with repeated measurement data, that is, when several measurements are taken on the same individual. Their approach consists of using a non-linear random parameter model and of defining prior distributions on the expected values and variance-covariance matrix of the random parameters. This seems to be a very attractive approach for estimating the parameters of agronomic models.

## REFERENCES

[1] Bennett J.E., Racine-Poon A., Wakefield J.C., MCMC for nonlinear hierarchical models, in: Gilks et al. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1995, pp. 339–357.

[2] Bernardo J.M., Smith A.F.M., Bayesian Theory, Wiley Chichester, 1994.

[3] Beven K., Binley A., The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process. 6 (1992) 279–298.

[4] Campbell E.P., Fox D.R., Bates B.C., A Bayesian approach to parameter estimation and pooling in nonlinear flood event models, Water Resour. Res. 35 (1999) 211–220.

[5] Fortran 90, Absoft Corporation, Rochester Hills, Michigan, 1998.

[6] Franks S.W., Gineste P., Beven K.J., Merot P., On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res. 34 (1998) 787–797.

[7] Gelfand A.E., Dey D.K., Chang H., Model determination using predictive distributions with implementation via sampling-based methods, in: Bernardo J.M., Berger J.O., Dawid A.P., Smith A.F.M. (Eds.), Bayesian Statistics, Vol. 4, Oxford University Press, New York, 1992, pp. 147–167.

[8] Gelman A., Inference and monitoring convergence, in: Gilks et al. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1995, pp. 131–142.

[9] Geyer C.J., Practical Markov chain Monte Carlo, Stat. Sci. 7 (1992) 473–511.

[10] Gilks W.R., Richardson S., Spiegelhalter D.J., Introducing Markov Chain Monte Carlo, in: Gilks et al. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1995, pp. 1–18.

[11] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[12] Harmon R., Challenor P., A Markov chain Monte Carlo method for estimation and assimilation into models, Ecol. Model. 101 (1997) 41–59.

[13] Makowski D., Wallach D., Meynard J.-M., Models of yield, grain protein, and residual mineral N reponses to applied N for winter wheat, Agron. J. 91 (1999) 377–385.

[14] Makowski D., Wallach D., Meynard J.-M., Statistical methods for predicting the responses to applied N and for calculating optimal N rates, Agron. J. 93 (2001) 531–539.

[15] Malakoff D., Bayes offers a ‘New’ way to make sense of numbers, Science 286 (1999) 1460–1464.

[16] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1091.

[17] Raftery A.E., Lewis S.M., Implementing MCMC, in: Gilks et al. (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1995, pp. 115–130.

[18] Shulz K., Beven K., Huwe B., Equifinality and the problem of robust calibration in nitrogen budget simulations, Soil Sci. Soc. Am. J. 63 (1999) 1934–1941.

[19] Wallach D., Regional optimization of fertilization using a hierarchical linear model, Biometrics 51 (1995) 338–346.