



**HAL**  
open science

## A nonlinear mixed effects model of plant growth and estimation via stochastic variants of the EM algorithm.

Charlotte Baey, Samis Trevezas, Paul-Henry Cournède

### ► To cite this version:

Charlotte Baey, Samis Trevezas, Paul-Henry Cournède. A nonlinear mixed effects model of plant growth and estimation via stochastic variants of the EM algorithm.. Communications in Statistics - Theory and Methods, 2015, pp.00. 10.1080/03610926.2014.930909 . hal-01068327

**HAL Id: hal-01068327**

**<https://hal.science/hal-01068327v1>**

Submitted on 25 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A nonlinear mixed effects model of plant growth and estimation via stochastic variants of the EM algorithm.

Charlotte Baey, Samis Trevezas\*, and Paul-Henry Cournède

Laboratory of Mathematics Applied to Systems,  
École Centrale Paris, Grande Voie des Vignes, 92290 Châtenay-Malabry, France

7 April 2014

## Abstract

There is a strong genetic variability among plants, even of the same variety, which, combined with the locally varying environmental conditions in a given field, can lead to the development of highly different neighboring plants. This is one of the reasons why population-based methods for modeling plant growth are of great interest. GreenLab is a functional structural plant growth model which has already been shown to be successful in describing plant growth dynamics primarily at individual level. In this study, we extend its formulation to the population level. In order to model the deviations from some fixed but unknown important biophysical and genetic parameters we introduce random effects. The resulting model can be cast into the framework of nonlinear mixed models, which can be seen as particular types of incomplete data models. A stochastic variant of an EM-type algorithm (Expectation-Maximization) is generally needed to perform maximum likelihood estimation for this type of models. Under some assumptions, the complete data distribution belongs to a subclass of the exponential family of distributions for which the M-step can be solved explicitly. In such cases, the interest is focused on the best approximation of the E-step by competing simulation methods. In this direction, we propose to compare two commonly used stochastic algorithms: the Monte-Carlo EM (MCEM) and the SAEM algorithm. The performances of both algorithms are compared on simulated data, and an application to real data from sugar beet plants is also given.

*Keywords:* plant growth model, nonlinear mixed effects model, stochastic EM algorithm, MCMC methods, sugar-beet plant

---

\*Corresponding author; e-mail: [samis.trevezas@ecp.fr](mailto:samis.trevezas@ecp.fr)

## 1 Introduction

There is a naturally high genetic variability among plants, even coming from the same variety, which is the guarantee for a better resistance to diseases or pest attacks, and enhance their ability to adapt to changing environmental conditions. In the same way, even in a given field, locally varying soil or environmental conditions can lead to the development of two significantly different neighboring plants. This inter-individual variability can have an impact at the agrosystem level, as shown for example by Brouwer et al. (1993), who demonstrated how soil and crop micro-variability can have an impact on the final yield, as some parts of the field can be more adapted to dryness and can compensate less good performances obtained in other parts of the field. Similarly, Renno and Winkel (1996) showed how asynchronous flowering can prevent plants from suffering from short-term stresses.

However, current practices in plant growth models rarely take this variability into account. Among the three different families of plant growth models (Vos et al., 2007), the most promising approach is the one based on functional-structural models (FSPM), describing the evolution of the plant structure, driven by the underlying ecophysiological processes (e.g. Sievänen et al. (2000)). These models make a fine description of the individual plant at the organ level and are an improvement over process-based models where plants are described only at the level of organ compartments and the crop production is only computed at the square meter level. However, while this level of description make these models very appealing, their calibration is often based on averaged individuals, which is not fully satisfactory since it only gives a partial representation of the field production, in addition to the subsequent loss of information. Moreover, the extrapolation of individual-based models to the field scale is still at its early stages, as far as plant growth is concerned. It mostly concerns competition for light (e.g. Cournède et al. (2008) or Fournier and Andrieu (1999)), and the growth of each individual plant is simulated taking into account a competition index. However, their application in real cases is difficult since it is not possible, in practice, to describe exhaustively all the individuals of a given population.

In this study, we propose an extension of the individual-based Greenlab model, based on a bottom-up approach: the growth of each individual plant in a given field can be characterized by the same set of equations as in the classical Greenlab model, but some of the model parameters are specific to this individual, and can therefore be considered as random effects. The resulting model can be cast into the framework of nonlinear mixed-effects models (Davidian and Giltinan, 1995, 2003), which can in turn be seen as incomplete data models, the random effects being the unobserved data. In this context, maximum likelihood estimators of the parameters can be obtained using an appropriate stochastic variant of the EM-algorithm (Expectation-Maximization) (Kuhn and Lavielle, 2005).

Due to the nonlinearity of the model, the E-step is in general analytically intractable, and an approximation of the so-called  $Q$ -function is therefore necessary. Nevertheless, under suitable assumptions, the M-step can be resolved explicitly. Moreover, in our case, the complete data distribution belongs to the

exponential family, for which simplifications of each step of the algorithm are available and convergence properties are established more easily. The observed Fisher Information Matrix can be easily computed in this case using Louis' missing information principle (Louis, 1982).

Several approaches have been developed to deal with an intractable E-step. A first approach was given by Wei and Tanner (1990) who proposed to replace exact evaluations of the  $Q$ -function by Monte-Carlo approximations. When direct simulation from the distribution of the hidden data given the observed ones is not possible, McCulloch (1994, 1997) suggested that a Markov Chain Monte Carlo approach can be used instead, which lead to the definition of the MCMC-EM algorithm. Other possibilities exist to simulate the hidden data, for example variants of sequential importance sampling and we refer the interested reader to Jank (2005) and Jank (2006a) for a general description of the different methods. In the context of individual plant growth modeling we also refer to Trevezas and Cournède (2013) and Trevezas et al. (2013) for adaptations and comparisons of the aforementioned stochastic variants. A different way of simulating the  $Q$ -function was originally proposed by Delyon et al. (1999). The authors used stochastic approximation (Robbins and Monro, 1951) of the  $Q$ -function (hence the name of SAEM for Stochastic Approximation of EM), where simulations of previous iterations are used with a weight that decreases with the distance to the current iteration. We propose in this paper a comparison of these two algorithms on several sets of simulated data, and an application to real data from sugar beet plants.

The rest of the paper is organized as follows. In Section 2, we introduce the Greenlab model and give its formulation as a nonlinear mixed model. In Section 3, we detail the maximum likelihood estimation using the EM algorithm and the computation of confidence intervals using the expected Fisher Information Matrix. The approximation of the  $Q$ -function using two stochastic variants of the EM-algorithm is given in Section 4: the MCMC-EM is described in Section 4.1 and the SAEM is described in Section 4.2. In Section 5 we apply the two aforementioned methods on simulated and real data sets from the sugar beet.

## 2 A population-based version of the Greenlab model

The Greenlab model is a functional-structural model, combining rules for (i) biomass (mass for living organisms) production and allocation (functional part), and (ii) architectural development at the organ level (structural part). It was introduced by de Reffye and Hu (2003), and represented as a discrete dynamic system in Cournède et al. (2006). Parameter estimation methods for this model are reviewed in Cournède et al. (2011). Some recent advances for parameter estimation in the presence of modelling errors can be found in Trevezas and Cournède (2013) and Trevezas et al. (2013). All these approaches assume that a single individual is observed, and when data from several plants are available, they rely on the construction of an 'average' plant. Here, we present how the individual-based model can be

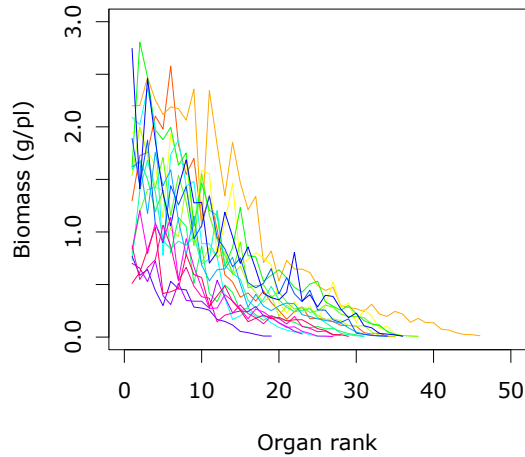


Figure 1: Individual blade biomasses according to the rank of the corresponding leaf. The curves for the individual petiole biomasses are of similar shape are not shown here for clarity.

extrapolated at the population scale to take into account the inter-individual variability, and we choose the sugar beet plant for our validation test. The sugar beet is a biennial plant that is usually harvested at the end of its first development stage for its fleshy root. At this stage, its architecture is quite simple as the plant can be decomposed into leaves (with two types of organs, blades and petioles) and the root system. We thus consider 3 types of organs denoted by  $\mathcal{O} = \{b, p, r\}$ , where  $b$  stands for blade,  $p$  for petiole and  $r$  for root. Each blade and each petiole are indexed by their rank  $n$  (starting from the inferior and older leaf). The data predicted by the model and from which the model hidden parameters will be estimated are illustrated in Figure 1: at the measurement time  $t_{max}$ , the masses of blades and petioles are given as functions their ranks for all the measured individuals.

## 2.1 Description of the Individual-Based Greenlab model

For an individual plant, the model predicts the evolution of all organ masses from germination, with a daily time step. The original version of the GreenLab model for sugar beet can be found in Lemaire et al. (2008). Three major biological phenomena have to be considered: organogenesis, that drives the creation of new organs, biomass production, by leaf photosynthesis, and biomass allocation, which corresponds to the distribution of the produced biomass to all the expanding and newly created organs according to their sink strengths (via a barycentric model).

**Organogenesis.** From the initial creation of the two cotyledons at germination, the sugar beet produces new leaves rhythmically. The main factor determining the rate of leaf appearance is temperature.

For this reason it is very convenient to introduce the notion of thermal time, which is defined as follows:

$$\tau(t) = \int_0^t \max(0, T(s) - T_b) ds, \quad t \geq 0,$$

and represents at calendar time  $t$ , the accumulated sum of temperatures above a base temperature  $T_b$  (chosen as  $0^\circ C$  for sugar beet, Lemaire et al. (2008)) from the time of germination until time  $t$ .

A piecewise linear function with two phases (corresponding to two distinct rates of leaf appearance) was shown to model pretty well sugar beet organogenesis Lemaire et al. (2008). In the sequel, for a leaf of rank  $n$ , we denote by  $\tau_n$  its thermal time of initiation,  $\tau_n^e$  its expansion period, and  $\tau_n^s$  its lifetime. The thermal time of initiation of root is thus equal to  $\tau_1$ , and we denote by  $\tau_r^e$  its corresponding expansion period. We assume that root does not get senescent, and that initiation, expansion and lifetime of the blade and petiole of the same leaf are identical.

We consider in this study that all the parameters of the organogenesis model ( $\tau_n, \tau_n^e, \tau_n^s$ , for all  $n$ ) are known (obtained by direct monitoring of the appearance, expansion and senescence of leaves, see Lemaire et al. (2008)).

**Biomass production.** The seed mass gives the initial biomass. After the appearance of the first couple of leaves (these leaves are called the cotyledons and are preformed in the seed), biomass production during day  $t$  is assumed to be given by:

$$F(t; p^*) = u_t \mu s^{pr} \left( 1 - \exp \left( -0.7 \frac{s^{act}(t; p^{al})}{s^{pr}} \right) \right), \quad t \geq 1, \quad (1)$$

where  $u_t$  stands for the environmental condition on day  $t$  (usually, the photosynthetically active radiation), 0.7 is the value of the Beer-Lambert coefficient of light extinction for the sugar beet,  $s^{pr}$  is an empirical coefficient related to the space occupied by the plant on the ground and is thus important to account for the competition between plants, specifically regarding light interception,  $\mu$  is the radiation use efficiency which is characteristic of the plant photosynthetic capacity,  $s^{act}(t; p^{al})$  is the photosynthetically active blade area at the beginning of day  $t$ . It is given by the sum of the surface areas of all non senescent blades, while the surface area of one blade is simply deduced from its mass, divided by a constant parameter  $e$  (mass per unit area, directly measured on plants). Therefore,  $s^{act}$  is written as a function of the allocation parameter vector  $p^{al}$  described in the next paragraph, and its equation is given in Equation (3). The set of unknown parameters is  $p^* = (\mu, s^{pr}, p^{al})$ .

**Biomass allocation.** A basic assumption of the Greenlab model is that biomass allocation to all expanding organs is proportional to organ specific functions, called sink functions. At a given time  $t$ , these functions depend on the type of organ and on its expansion stage

The initial phase corresponds to germination (from initiation to emergence) in which the seed biomass is distributed to the cotyledons and root system. After emergence, the biomass produced by photosynthesis given by (1) is distributed to all expanding organs. In both cases, at time  $t$ , allocation to an organ of type  $o$  and of rank  $n$ , is proportional to the following sink strength:

$$s_{o,n}(t; p_o^{al}) = c \gamma_o \left( \frac{\tau(t) - \tau_n}{\tau_n^e} \right)^{a_o - 1} \left( 1 - \frac{\tau(t) - \tau_n}{\tau_n^e} \right)^{b_o - 1} \mathbf{1}_{\tau_n \leq \tau(t) \leq \tau_n + \tau_n^e},$$

where  $p_o^{al} = (\gamma_o, a_o, b_o)$  for  $o \in \mathcal{O}$ ,  $c$  is the normalizing constant of the beta law  $B(a_o, b_o)$  and  $\gamma_o$  is a multiplicative factor relative to the biomass attraction strength for the organ of type  $o$ . By convention, and since it is a barycentric model, we choose  $\gamma_b = 1$  for the blades.

The sum of all sink strengths on day  $t$  defines the total biomass demand  $d(t; p^{al})$ . It is simply given by:

$$d(t; p^{al}) = s_{r,1}(t; p_r^{al}) + \sum_{o=b, p} \sum_{n \geq 1} s_{o,n}(t; p_o^{al}).$$

Note that in the sum, most sinks are null (only those for organs whose initiation time and expansion duration has not elapsed are non null). The ratio  $s_{o,n}(t; p_o^{al})/d(t; p^{al})$  determines the percentage of the produced biomass  $F(t; p^*)$  which is allocated to the organ of type  $o$  and rank  $n$  at the end of day  $t$ . We can thus deduce  $q_{o,n}(t; p^{al})$ , the mass of organ of type  $o$  and rank  $n$  at the beginning of day  $t$ :

$$q_{o,n}(t; p^*) = \sum_{u=1}^{t-1} s_{o,n}(u; p_o^{al}) F(u; p^*) / d(u; p^{al}). \quad (2)$$

Finally, the active blade surface area at time  $t$  is given by:

$$s^{act}(t; p^{al}) = \frac{1}{e} \left( \sum_{n \in \mathcal{L}(t)} \sum_{u=1}^{t-1} s_{b,n}(u; p_b^{al}) F(u; p^*) / d(u; p^{al}) \right) \quad (3)$$

with  $\mathcal{L}(t) = \{n \geq 1 \text{ such that } \tau(t-1) - \tau_n < \tau_n^s\}$ , the set of all non-senescent leaves at the beginning of day  $t$ .

## 2.2 A two-stage formulation of the model

In order to account for inter-individual variability, random effects are introduced in the Greenlab model, which can then be seen as a two-stage hierarchical one. In the first stage we characterize the intra-individual variability, i.e., the variation of measurements from the same individual, using the Greenlab model. We obtain for each plant a set of subject-specific parameters, the inter-individual variation of which is accounted for in a second stage.

**First-stage: intra-individual variation.**

Let us denote by  $y = (y_{i,n})_{1 \leq i \leq s, 1 \leq n \leq n_i}$  the log-values of the observed biomasses of organs of rank  $n$  for plant  $i$ , at the beginning of day  $t_{max}$  (corresponding to the measurement day), with  $n_i$  the number of leaves for plant  $i$ . By defining  $\log x := (\log x_1, \dots, \log x_d)$  for a  $d$ -dimensional vector  $x$ , we assume the following log-additive measurement error model:

$$\begin{aligned} y_{i,n} &= \log G_n(\phi_i) + \epsilon_{i,n}, \\ \epsilon_{i,n} &\sim \mathcal{N}_{d_n}(0, \Sigma_n), \end{aligned} \quad (4)$$

where  $\phi_i$  is the vector of parameters specific to plant  $i$ ,  $G_n$  is the vector-valued function of the theoretical biomasses of organs of rank  $n$  deduced from Equation (2) and given by:

$$G_n(\phi_i) = \left( \sum_{t=1}^{t_{max}-1} \frac{s_{o,n}(t; p^{al})}{d(t; p^{al})} F(t; \phi_i) \right)_{o \in \mathcal{O}}, \quad (5)$$

$t_{max}$  is the day at which observations are made and all the other quantities which appear in (5) are defined in Section 2.1. The error terms  $(\epsilon_{i,n})_{1 \leq i \leq s, 1 \leq n \leq n_i}$  are mutually independent random variables and

$$\Sigma_n = \begin{cases} \text{diag}\{\Sigma_{b,p}, \sigma_r^2\}, & \text{if } n = 1, \\ \Sigma_{b,p}, & \text{if } n > 1, \end{cases} \quad (6)$$

where  $\Sigma_{b,p}$  is a  $2 \times 2$  covariance matrix for blade and petiole measurement errors and  $\sigma_r^2$  is the variance of the root measurement error. We thus make the assumptions that measurement errors from organs of two different plants, or of the same plant but with different ranks are independent.

**Second-stage: inter-individual variation.**

In this second stage we assess the variability of the subject-specific parameters  $\phi_i$ . We assume the following model for the vector  $\phi_i = (\phi_{i,1}, \dots, \phi_{i,P})^t$ , with  $P$  the number of random parameters:

$$\begin{aligned} \phi_i &= \beta + \xi_i, \\ \xi_i &\sim \mathcal{N}_P(0, \Gamma), \end{aligned} \quad (7)$$

where  $\beta$  is the  $P$ -dimensional vector of fixed effects and  $\Gamma = \text{diag}\{\sigma_1^2, \dots, \sigma_P^2\}$  is a diagonal covariance matrix. If the domain for some of the parameters under consideration is not  $\mathbb{R}$ , either due their definition or because of biological constraints, then it could be possible to make appropriate reparameterizations in such a way that (7) holds for the transformed parameters. Nevertheless, we stress the fact that multinormal assumptions could also not be valid in some application contexts. Moreover, the number of parameters which are considered with random effects could be a subset of  $p^*$  given in Section 2.1. For



the specific choice that we make in the simulated example and the real data application, see Section 5. For some possibilities to take into account fixed effects we refer to Remark 3 and the comments before.

### 3 Maximum likelihood estimation

Let us denote the overall parameter vector by  $\theta$ , with  $\theta \in \Theta \subseteq \mathbb{R}^d$ . In our case,  $\theta = (\beta, \sigma_1^2, \dots, \sigma_P^2, \Sigma_{b,p}, \sigma_r^2)^t$  and it contains the parameters which are used to describe the intra-individual and the inter-individual variability as well as those related to the distribution of the measurement errors. In order to compute the maximum likelihood estimator (MLE) of  $\theta$ , we need to compute the likelihood of the model, which will be in general analytically intractable due to the nonlinearity of  $G_n$ . However, our model can be seen as an incomplete data model, with  $y = (y_{i,n}, 1 \leq i \leq s, 1 \leq n \leq n_i)$ , the observed data, and the random effects  $\phi = (\phi_i, 1 \leq i \leq s)$  being the unobserved data.

In the sequel, we will denote each probability density by  $f(\cdot; \theta)$ , and the associated random variables will be distinguished by the first argument of  $f$ . For example,  $f(y; \theta)$  denotes the probability density of the observed data,  $f(\phi; \theta)$  the density of the unobserved data or random effects, and  $f(y | \phi; \theta)$  the density of the observed data given the unobserved ones. The likelihood of the observed data  $y$  can then be expressed according to the likelihood of the complete data  $x := (y, \phi)$  using:

$$L(y; \theta) := \int f(y, \phi; \theta) d\phi = \int f(y | \phi; \theta) f(\phi; \theta) d\phi. \quad (8)$$

In such cases, an appropriate variant of an EM-algorithm (Expectation-Maximization) can be implemented to approximate the MLE. The EM-algorithm has a long history: the underlying principles of the algorithm were first evoked by Orchard and Woodbury (1972) as the ‘‘missing information principle’’ and by Sundberg (1974) in the special case of the exponential family. Dempster et al. (1977) presented a generalization of the algorithm and named it the EM-algorithm. It was further developed by many authors to cover extensions and generalizations (see McLachlan and Krishnan (2007) for more details).

Each iteration of the EM algorithm consists in two steps: (i) an expectation step (E-step) in which the conditional expectation of the complete data log-likelihood given the observed data, the so-called  $Q$ -function, is computed under the current parameter value, and (ii) a maximization step (M-step) in which the parameters are updated by maximizing the  $Q$ -function obtained in the E-step. At each iteration, the increase of the  $Q$ -function results in an increase of the incomplete data log-likelihood.

However, in the special case where the complete data likelihood belongs to the exponential family, these two steps can be written in a simpler form with the use of sufficient statistics. Indeed, in this case the probability density of the complete data can be written as:

$$f(y, \phi; \theta) = h(y, \phi) \exp \{ \langle s(\theta), t(y, \phi) \rangle - a(\theta) \},$$

where  $\langle \cdot, \cdot \rangle$  denotes the scalar product on  $\mathbb{R}^n$ . It is more convenient to work with the natural parameterization  $\eta = s(\theta)$  of the model, with  $s : \Theta \mapsto \Omega$ , for which calculations are easier. If the dimensions of  $\Theta$  and  $\Omega$  are identical, the model is said to belong to the *regular* exponential family. The density can then be written as:

$$f(y, \phi; \eta) = h(y, \phi) \exp \{ \langle \eta, t(y, \phi) \rangle - b(\eta) \}, \quad (9)$$

where

$$e^{b(\eta)} = \int e^{\langle \eta, t(y, \phi) \rangle} h(y, \phi) d\phi dy. \quad (10)$$

In particular,  $b(\eta)$  is the logarithm of the Laplace transform of the measure  $\nu \circ t^{-1}$ , where  $\nu(d\phi, dy) := h(y, \phi)\lambda(d\phi, dy)$ , and  $\lambda$  is the Lebesgue measure of an appropriate dimension indicated by  $(y, \phi)$ . It is therefore analytic and by differentiation under the integral sign, we have the following relationship:

$$\nabla b(\eta) = \mathbb{E}_\eta(t(y, \phi)). \quad (11)$$

Another interesting property of the exponential family is the relationship between  $b$ , the sufficient statistics and the Fisher information matrix given by:

$$\text{Cov}_\eta(t(y, \phi)) = \nabla^2 b(\eta) = \mathcal{I}(\eta). \quad (12)$$

In our case, let us decompose  $\theta$  as  $\theta = (\theta_1, \theta_2)$ , where  $\theta_1 = (\beta, \Gamma)$  and  $\theta_2 = (\Sigma_{b,p}, \sigma_r^2)$ . By the particular structure of this model and the multinormal assumptions given in Section 2.2, we have

$$f(y, \phi; \eta) = \exp \{ \langle \eta_1, t_1(\phi) \rangle - b_1(\eta_1) \} \exp \{ \langle \eta_2, t_2(y, \phi) \rangle - b_2(\eta_2) \}, \quad (13)$$

where

$$\begin{aligned} \eta_1 &= s_1(\theta_1) = (\Gamma^{-1}\beta, \Gamma^{-1})^t := (\eta_{11}, \eta_{12})^t, \\ \eta_2 &= s_2(\theta_2) = (\Sigma_{b,p}^{-1}, \sigma_r^{-2})^t := (\eta_{21}, \eta_{22})^t, \\ t_1(\phi) &= \left( \sum_{i=1}^s \phi_i, -\frac{1}{2} \sum_{i=1}^s \phi_i \phi_i^t \right)^t, \\ t_2(y, \phi) &= -\frac{1}{2} \left( \sum_{i=1}^s \sum_{n=1}^{n_i} (y_{i,n} - G_n(\phi_i)) (y_{i,n} - G_n(\phi_i))^t, \sum_{i=1}^s (y_{i,1}^r - G_1^r(\phi_i))^2 \right)^t, \\ b_1(\eta_1) &= \frac{sP}{2} \log 2\pi - \frac{s}{2} \log |\eta_{12}| + \frac{s}{2} \eta_{11}^t \eta_{12}^{-1} \eta_{11}, \\ b_2(\eta_2) &= \frac{s + 2 \sum_{i=1}^s n_i}{2} \log 2\pi - \frac{\sum_{i=1}^s n_i}{2} \log |\eta_{21}| - \frac{s}{2} \log \eta_{22}. \end{aligned} \quad (14)$$

**Remark 1.** *The use of matrices for the sufficient statistics leads to the use of the following scalar product*

on  $\mathcal{M}_m(\mathbb{R})$ : for every matrices  $A$  and  $B$  in  $\mathcal{M}_m(\mathbb{R})$ ,  $\langle A, B \rangle := \text{Tr}(A^t B)$ .

The two steps of the EM-algorithm are described below:

### 3.1 E-step

At iteration  $k$ , the E-step of the algorithm consists in the computation of the  $Q$ -function given the current parameter value  $\theta^k$ . In the case where the complete data distribution belongs to the exponential family, this step is reduced to the computation of the conditional expectation of the sufficient statistics given the observed data  $y$  under the current parameter value (see, e.g., McLachlan and Krishnan (2007), p. 22):

$$t^{(k)} = \mathbb{E}_{\theta^k}(t(y, \phi) \mid y), \quad (16)$$

and the  $Q$ -function is given by  $Q(\theta; \theta^k) = \langle s(\theta), t^{(k)} \rangle - a(\theta)$ . In our case, by (13), (15) and our assumptions, the step is reduced to the computations

$$t_1^{(k)} = \mathbb{E}_{\theta^k}(t_1(\phi) \mid y) = \sum_{i=1}^s \mathbb{E}_{\theta^k}(t_{i,1}(\phi_i) \mid y_i), \quad (17)$$

$$t_2^{(k)} = \mathbb{E}_{\theta^k}(t_2(y, \phi) \mid y) = \sum_{i=1}^s \mathbb{E}_{\theta^k}(t_{i,2}(y_i, \phi_i) \mid y_i), \quad (18)$$

where  $t_{i,1}$  and  $t_{i,2}$  correspond to the  $i$ -th summand of  $t_1$  and  $t_2$  respectively, given by (15).

Unfortunately, these conditional expectations can not be computed analytically, and have to be approximated. We discuss in Section 4 two approximation methods based on Markov Chain Monte Carlo methods.

### 3.2 M-step

In the M-step of the algorithm, we maximize the  $Q$ -function with respect to (w.r.t.)  $\theta$ . In the exponential family case, this step reduces to the resolution of the following equation:

$$\nabla_{\eta} b(\eta) = t^{(k)}. \quad (19)$$

This can be seen directly by taking the conditional expectation of the logarithms of each member of (9) (under  $\theta^k$ ) and by setting the gradient w.r.t.  $\eta$  equal to zero. In several cases this equation can be solved explicitly and this leads to an explicit M-step. In our case the following result holds.

**Proposition 1.** *Under the assumption that the covariance matrix  $\Gamma = \text{diag}\{\sigma_1^2, \dots, \sigma_p^2\}$  of the random effects is of full rank, then the EM-update equations of the nonlinear mixed model given by (4) and (7)*

are given as follows:

$$\begin{aligned}
\hat{\beta} &= \frac{1}{s} \sum_{i=1}^s \mathbb{E}_{\theta^k}(\phi_i | y_i), \\
\hat{\sigma}_j^2 &= \frac{1}{s} \sum_{i=1}^s \mathbb{E}_{\theta^k}(\phi_{i,j}^2 | y_i) - \hat{\beta}_j^2, \quad j = 1, \dots, P, \\
\hat{\Sigma}_{b,p} &= \frac{1}{\sum_{i=1}^s n_i} \sum_{i=1}^s \sum_{n=1}^{n_i} \mathbb{E}_{\theta^k} \left[ (y_{i,n} - \log G_n(\phi_i)) (y_{i,n} - \log G_n(\phi_i))^t | y_i \right], \\
\hat{\sigma}_r^2 &= \frac{1}{s} \sum_{i=1}^s \mathbb{E}_{\theta^k} \left[ (y_{i,1}^r - \log G_1^r(\phi_i))^2 | y_i \right], \tag{20}
\end{aligned}$$

where  $y_{i,1}$  in the expression of  $\hat{\Sigma}_{b,p}$  is restricted to blade and petiole only.

*Proof.* Due to the particular decomposition of  $f(y, \phi; \eta)$  given by (13), solving equation (19), which is generally valid for members of the exponential family, is equivalent to solving separately for  $j = 1, 2$ ,

$$\nabla_{\eta_j} b_j(\eta_j) = t_j^{(k)}, \tag{21}$$

w.r.t.  $\eta_j$ , where  $\eta_j$  and  $b_j(\eta_j)$  are given by (15) and  $t_j^{(k)}$  by (17). It is easy to obtain the solutions of (21) and then use the inverse transformations  $s_i^{-1}(\eta_i)$  to obtain the solutions for  $\theta_1$  and  $\theta_2$ . The solutions for all the components are given by (20) and indeed correspond to a unique maximum.  $\square$

We can see from equations (11), (16) and (19) that in the exponential family case, the repetition of the two steps of the algorithm amounts to iteratively fit  $\theta$  by trying to match the conditional and the unconditional expectations: ultimately, when the algorithm has reached convergence for a given  $\tilde{\theta}$ , these two quantities are indeed equal.

### 3.3 Convergence of the EM algorithm

In the special case where the complete-data likelihood belongs to the exponential family, the convergence of the EM algorithm has been studied by various authors, including Sundberg (1974), Dempster et al. (1977), and Wu (1983), which corrected a mistake in the proof of Dempster et al. (1977) (see also Delyon et al. (1999) or Fort and Moulines (2003)). We recall here the assumptions required for the convergence of the EM as given by Fort and Moulines (2003) and show how they apply in our case.

In the sequel,  $\lambda$  will denote the Lebesgue measure on  $\mathbb{R}^n$ , and  $\ell_c$  the log-likelihood of the complete data, so that  $\ell_c(t; \theta) := \langle s(\theta), t \rangle - a(\theta)$ . The required assumptions are given as follows:

(M1) The parameter space  $\Theta$  is an open subset of  $\mathbb{R}^d$ , with  $d$  the dimension of  $\theta$ .

(M2) (a) The functions  $a : \Theta \rightarrow \mathbb{R}$  and  $s : \Theta \rightarrow \mathbb{R}^d$ , defined in (13), are continuous on  $\Theta$ , and  $t : \mathbb{R}^{q+l} \rightarrow \mathcal{T} \subseteq \mathbb{R}^d$  is continuous on  $\mathbb{R}^l$ ,

- (b) for all  $\theta$  in  $\Theta$ ,  $\bar{t}(\theta) := \int t(y, \phi) f(\phi|y; \theta) \lambda(d\phi)$  is finite and continuous on  $\Theta$ ,
- (c) there exists a continuous function  $\hat{\theta} : \mathcal{T} \rightarrow \Theta$  such that for all  $t \in \mathcal{T}$ ,  $\ell_c(t; \hat{\theta}(t)) = \sup_{\theta \in \Theta} \ell_c(t, \theta)$ ,
- (d) the observed-data likelihood  $L$  defined by  $L(\theta) = \int f(y, \phi; \theta) \lambda(d\phi)$  is positive, finite and continuous on  $\Theta$ ,
- (e) for all  $M > 0$ , the level set  $\{\theta \in \Theta, L(\theta) \geq M\}$  is compact.

(M3) Denoting the set of stationary points of the algorithm by  $\mathcal{L} := \{\theta \in \Theta, \hat{\theta} \circ \bar{t}(\theta) = \theta\}$ , we have that either  $L(\mathcal{L})$  is compact or for all compact sets  $\mathcal{K} \subseteq \Theta$ ,  $L(\mathcal{L} \cap \mathcal{K})$  is finite.

**Remark 2.** *i) Assumptions (M2)(a) and (b) correspond to the assumption made by Wu (1983) that the Q-function must be continuous on both  $\theta$  and  $\theta^k$ . Indeed, in the case of a regular exponential family, the Q-function is always continuous on  $\theta$  whenever  $s$  and  $a$  are continuous, which is controlled by (M2)(a), and it is continuous on  $\theta^k$  when (M2)(b) is valid.*

*ii) Assumption (M2)(e) could be restrictive for applications. Nevertheless, it could be weakened as follows:*

*(e\*) there exists an  $M_0 \geq 0$ , such that for all  $M > M_0$ , the level set  $\{\theta \in \Theta, L(\theta) \geq M\}$  is compact.*

*Notice that for  $M_0 = 0$ , (M2)(e\*) is identical to (M2)(e), so (M2)(e\*) is weaker than (M2)(e). Moreover, the ascent property of the EM algorithm guarantees that if there exists an  $n \in \mathbb{N}$ , such that  $L(\theta^n) > M_0$ , then under this condition the sequence of EM iterates finally concentrates on a compact subset.*

In our case, as we assumed a diagonal covariance matrix for the random effects, we have the following representation of  $\theta = (\beta, \sigma^2, \sigma_b^2, \sigma_p^2, \rho, \sigma_r^2)$ , where  $\sigma^2$  is a vector of size  $P$  containing the variances of the random effects (i.e. the diagonal elements of  $\Gamma$ ) and we have:

- $\beta \in \mathbb{R}^P$  for  $j = 1, \dots, P$  (transformations can be applied to the original parameters to ensure that their domain of variation is  $\mathbb{R}$ , for example logarithmic transformation for strictly positive parameters)
- $\sigma^2 \in (\mathbb{R}_+^*)^P$ ,  $\sigma_b^2, \sigma_p^2, \sigma_r^2 \in \mathbb{R}_+^*$ ,
- $\rho \in (-1, 1)$ , which corresponds to the assumption that the covariance matrix  $\Sigma_{b,p}$  is non-singular (given that  $\sigma_b^2, \sigma_p^2 \in \mathbb{R}_+^*$ ).

Finally,  $\Theta = \mathbb{R}^P \times (\mathbb{R}_+^*)^{P+3} \times (-1, 1)$  which is an open subset of  $\mathbb{R}^{2P+4}$ , so (M1) is verified.

(M2)(a) is trivially verified, and assumption (M2)(c) is also valid using the explicit equations for the M-step given by (20). Indeed, the function  $\theta \mapsto \ell_c(t; \theta)$  is strictly concave so that  $\hat{\theta}$  is well defined, and it is also continuous as a consequence of the implicit function theorem. Assumption (M2)(b) follows from the properties of the exponential family, as explained in Sundberg (1974). The conditional probability

density of  $\phi$  given  $y$  w.r.t. the Lebesgue measure on  $\mathbb{R}^l$  and using the natural parametrization  $\eta = s(\theta)$  is given by:

$$f(\phi|y; \eta) = \exp\{\langle \eta, t(y, \phi) \rangle - b_y(\eta)\}, \quad (22)$$

where

$$e^{b_y(\eta)} = \int e^{\langle \eta, t(y, \phi) \rangle} \lambda(d\phi). \quad (23)$$

Similarly to the function  $b$  defined in (10),  $b_y$  is also the Laplace transform associated with the image measure of  $\lambda$  by the application  $t$ , from which it follows that it is infinitely differentiable on  $\Omega$ . As we have:

$$\nabla b_y(\eta) = \mathbb{E}_\eta(t(y, \phi) | y) = \bar{t}(\eta), \quad (24)$$

the application  $\bar{t}$  is continuous and finite on  $\Omega$ . As the inverse transformation  $s^{-1}(\eta)$  is continuous,  $\bar{t}$  is also continuous and finite on  $\Theta$  and (M2)(b) is verified.

Condition (M2)(d) also holds. By (9) with  $h = 1$  (see also (13)) and (22), we finally have that  $L(\eta) = \exp\{b_y(\eta) - b(\eta)\}$ , where  $b_y(\eta)$  and  $b(\eta)$  are analytic and this shows that  $L$  is positive, finite and continuous.

In the following proposition we give a very weak condition to ensure that  $L(\theta)$  is bounded, which is of course a prerequisite for the existence of the MLE. The proof can be found in the Appendix, together with a second proof of the continuity of  $L$  without using Laplace transforms. This proof has an independent interest and uses a different representation of the likelihood function which is particularly adapted to models with random effects and it also illustrates that condition (M2)(e\*) that we give in this paper is more appropriate than (M2)(e), since the latter is generally not satisfied for models with random effects.

**Proposition 2.** *The likelihood function  $L(\theta)$  is continuous and if we assume that this condition holds:  $\inf_{\beta \in \mathbb{R}^P} (\det \Sigma(\beta)) > 0$  and also  $\inf_{\beta \in \mathbb{R}^P} (\sigma_r^2(\beta)) > 0$ , then  $L(\theta)$  is also bounded.*

By the previous arguments it is clear that (M2)(e) is not satisfied since as the variance of a random effect tends to zero, then the resulting model is the one that we have by considering this parameter as a fixed effect, and consequently the level set  $\{\theta \in \Theta, L(\theta) \geq M\}$  cannot be compact. Nevertheless, condition (M2)(e\*) can still be verified. For example, in the case that all parameters that approach the boundary except for the variances of the random effects give  $L(\theta) \rightarrow 0$  and  $\sup_j L(\theta_{-j}, \beta_j) < \sup_\theta L(\theta)$ . In practice, however, if the variances of some random parameters tend to 0, the corresponding effects will be considered as fixed effects and will be incorporated in  $\theta_2$ . In this case, several cases can arise: if the MLE of the associated fixed effects can be expressed explicitly, we proceed as stated before. If no explicit maximization is possible, we can check whether conditionally on all the other parameters, an explicit solution can be derived, in which case a conditional version of the EM algorithm, ECM, can be used. If it is still not possible to obtain an explicit expression for the update equations of these parameters, then

alternatives such as quasi-Newton methods can be used (see Trevezas and Cournède (2013); Trevezas et al. (2013) for application of these methods in the case of the Greenlab model).

**Remark 3.** *It is also possible, as noted by Kuhn and Lavielle (2005) and first proposed by Racine-Poon (1985), to consider a Bayesian formulation for the fixed effects, that is to say, to assume prior distributions for these parameters, and to take for example the posterior mean or the posterior mode as estimates. Indeed, the use of a Bayesian formulation allows for the model to be cast again in the exponential family, but on the other hand, we are no longer in the framework of maximum likelihood estimation. Moreover, care must be taken in the choice of the prior distribution.*

$\mathcal{L}$  is here clearly compact, so that (M3) holds.

### 3.4 Confidence intervals

Let  $\theta^*$  denote the true unknown value of parameter  $\theta$ , and  $\hat{\theta}$  the estimator obtained with the EM algorithm. In the special case of the exponential family, Sundberg (1974) showed that under what he called the “ $n^{1/2}$ -consistency condition” which will be detailed below,  $\hat{\theta}$  is a consistent estimator of  $\theta^*$ , and this estimator is unique with a probability tending to one in a neighbourhood of  $\theta^*$ . Moreover,  $\hat{\theta}$  is asymptotically efficient and asymptotically normal with:

$$\sqrt{s} \left( \hat{\theta} - \theta^* \right) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}(\theta^*)^{-1}), \quad (25)$$

where  $\mathcal{I}(\theta^*)$  is the (expected) Fisher Information Matrix (FIM).

The  $n^{1/2}$ -consistency condition in Sundberg (1974) states that the matrix  $\mathcal{I}(\theta^*)$  must be strictly positive definite. In the case of the exponential family which we are dealing with, this assumption is sufficient to ensure the consistency and asymptotic properties of the MLE (see Sundberg (1974)).

To build asymptotic confidence intervals for  $\theta$ , we can use the consistent estimator of the FIM,  $\mathcal{I}(\hat{\theta})$ . In our context however, this matrix cannot be computed in a closed-form as it depends on the unknown incomplete data density, and we will use the missing information principle introduced by (Orchard and Woodbury, 1972) to approximate it Louis (1982).

Let us first define the *observed* information matrix as  $I(\theta; y) = -\nabla_{\theta}^2 \log f(y; \theta)$  and the relationship between this matrix and the *expected* information matrix is as follow:  $\mathcal{I}(\theta) = \mathbb{E}_{\theta}(I(\theta; y) \mid y)$ . Efron and Hinkley (1978) showed that in the case of an unidimensional parameter, the inverse of the observed information is a better approximation of the asymptotic variance of the maximum likelihood estimator than the inverse of the expected information, when evaluated at  $\hat{\theta}$ . It is also much simpler to compute. We have the following relationship (we refer the reader to McLachlan and Krishnan (2007) for more

details):

$$I(\theta, y) = \mathcal{I}_c(\theta, y) - \mathcal{I}_m(\theta, y) \quad (26)$$

$$:= \mathbb{E}_\theta [I_c(\theta, y) | y] - \mathbb{E}_\theta [S_c(x; \theta)S_c(x; \theta)^t | y], \quad (27)$$

where  $S_c(x; \theta) = \nabla_\theta \log f(y, \phi; \theta)$ . This is an application of the “missing information principle” introduced by Orchard and Woodbury (1972): the observed information corresponds to the complete information minus the missing information.

In the next proposition we recall the simplifications that can be made if the case of the exponential family.

**Proposition 3.** *In the case of a regular exponential family, the complete-data information matrix and the missing information matrix are given as follows:*

$$\mathcal{I}_c(\theta; y) = \text{Cov}_\theta(t(x)), \quad (28)$$

$$\mathcal{I}_m(\theta; y) = \text{Cov}_\theta(t(x) | y), \quad (29)$$

where  $t(x)$  is the vector of sufficient statistics defined in (14).

*Proof.* See Sundberg (1974) or McLachlan and Krishnan (2007). The first matrix does not depend on the data and can be expressed as:

$$\mathcal{I}_c(\theta; y) = \mathcal{I}_c(\theta) = \text{Cov}_\theta(t(x)) = J_\eta(\theta)^t \nabla_{\eta(\theta)}^2 a(\theta) J_\eta(\theta), \quad (30)$$

where  $J_\eta$  is the Jacobian matrix of  $\eta$  and  $\nabla^2 a$  the Hessian matrix of  $a$ . □

In the next corollary we give the form of the complete-data information matrix in our case.

**Corollary 1.** *The complete-data information matrix of the non linear mixed model given by (4) and (7) is given in a block-diagonal form as*

$$\mathcal{I}_c(\theta) = \text{diag} \left\{ \mathcal{I}_c(\theta_{1,j})_{1 \leq j \leq P}, \mathcal{I}_c(\theta_{2,1}), \frac{s}{2\sigma_r^4} \right\} \quad (31)$$

where

$$\mathcal{I}_c(\theta_{1,j}) = \begin{pmatrix} \frac{s}{\sigma_j^2} & 0 \\ 0 & \frac{s}{2\sigma_j^4} \end{pmatrix}, \quad j = 1, \dots, P, \quad (32)$$



and where

$$\mathcal{I}_c(\theta_{2,1}) = \frac{N}{1 - \rho^2} \begin{pmatrix} \frac{(2 - \rho^2)}{4\sigma_b^4} & -\frac{\rho^2}{4\sigma_b^2\sigma_p^2} & -\frac{\rho}{2\sigma_b^2} \\ -\frac{\rho^2}{4\sigma_b^2\sigma_p^2} & \frac{(2 - \rho^2)}{4\sigma_p^4} & -\frac{\rho}{2\sigma_p^2} \\ -\frac{\rho}{2\sigma_b^2} & -\frac{\rho}{2\sigma_p^2} & \frac{1 + \rho^2}{1 - \rho^2} \end{pmatrix} \quad (33)$$

A consistent estimator of this matrix can be obtained using  $\mathcal{I}_c(\hat{\theta})$ . On the other hand, the computation of the second matrix  $\mathcal{I}_m(\theta, y)$  involves, as for the non-explicit E-step, computations of conditional expectations that can not be done explicitly. Thus, it will be approximated by means of an appropriate stochastic variant of the EM algorithm.

The confidence intervals obtained thanks to this approach will be compared to those obtained via parametric bootstrap.

## 4 Approximation of the E-step

The E-step given by (17) involves computations from the conditional distributions of the hidden data  $\phi_i$  given the observed data  $y_i$ , under the current value of the parameter  $\theta^k$ , denoted by  $f(\phi_i | y_i; \theta^k)$ . Unfortunately, due to the nonlinearity of the model, these computations can not be done explicitly. However, many stochastic variants of the EM algorithm are available to approximate a non-explicit E-step. In this section, we discuss two algorithms based on Markov Chain Monte Carlo simulations. The first one is the Monte Carlo-EM algorithm, introduced by Wei and Tanner (1990) for the Monte Carlo part, and further extended by McCulloch (1994, 1997) to include Markov Chain Monte Carlo Simulations. The second algorithm presented in this paper is the SAEM algorithm Delyon et al. (1999), where a stochastic approximation (Robbins and Monro, 1951) is done at each iteration, and that has also been extended to include Markov Chain Monte Carlo simulations by Kuhn and Lavielle (2004).

Convergence of the MCMC-EM and the SAEM algorithm (under mild assumptions) towards a (local) maximum of the observed-data likelihood has been shown by Fort and Moulines (2003) and Delyon et al. (1999), respectively. The required assumptions will be recalled in Sections 4.1.2 and Section 4.2.2 respectively.

Once we are able to simulate from the conditional distribution of the hidden data given the observed data, under the current parameter value, the M-step simply consists in replacing the conditional expectations with the corresponding ergodic means.

## 4.1 The MCMC-EM algorithm

### 4.1.1 Description

When the conditional expectations involved in the E-step can not be evaluated analytically, Wei and Tanner (1990) suggested to approximate these quantities by Monte Carlo simulations. However, when direct simulation from  $f(\phi_i | y_i; \theta^k)$  is impossible, McCulloch (1994, 1997) proposed to use Markov Chain Monte Carlo (MCMC) approaches instead. More specifically, at each iteration  $k$  of the algorithm, a Markov Chain of size  $m_k$ , with stationary distribution  $f(\phi_i | y_i; \theta^k)$  is generated for each plant  $i$ . If we denote by  $(\phi_i^{k,(1)}, \dots, \phi_i^{k,(m_k)})$  the  $m_k$  realizations of the MCMC procedure for plant  $i$ , at iteration  $k$  and after a burn-in period, the E-step can then be approximated by :

$$t_1^{(k)} = \sum_{i=1}^s \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_{i,1}(\phi_i^{k,(m)}) \right), \quad (34)$$

$$t_2^{(k)} = \sum_{i=1}^s \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_{i,2}(y_i, \phi_i^{k,(m)}) \right). \quad (35)$$

We can also approximate the missing information matrix  $\mathcal{I}_m(\hat{\theta})$ :

$$\hat{\mathcal{I}}_m^{(k)}(\theta; y) = \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} t_m^{(k)t} - \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} \right) \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} \right)^t. \quad (36)$$

### 4.1.2 Convergence issues

Assumptions (M1)-(M3) of Section 3.3 concern the convergence of the EM algorithm. For the convergence of the stochastic variant of the algorithm MCMC-EM, it is necessary to add more assumptions regarding the Markov Chain Monte Carlo process. In their paper, Fort and Moulines (2003) introduce a *stable* version of the algorithm, which consists in the definition of a sequence of compact subsets  $(\mathcal{K}_n)$  where  $\mathcal{K}_n \subsetneq \mathcal{K}_{n+1}$  and  $\Theta = \bigcup_n \mathcal{K}_n$ , such that at each iteration  $k+1$  of the MCMC-EM algorithm, if the parameter updates (as defined in (20)) fall outside of the compact set  $\mathcal{K}_{p_k}$ , we reinitialize the parameters with  $\theta^{k+1} = \theta^0$ , and let  $p_{k+1} = p_k + 1$  where  $p_k$  counts for the number of reinitializations.

Using this stable version of the MCMC-EM algorithm, the authors add an assumption concerning the  $L^p$ -norm of the fluctuations of the MCMC approximation made in (20). In practice, the given hypothesis is always valid when the transition kernel of the MCMC algorithm is uniformly ergodic, which will be the case for example for the independent Metropolis-Hastings algorithm, or for some adaptive MH algorithms (as shown for example by Andrieu and Moulines (2006)). A second assumption concerns the Markov Chain sizes  $\{m_k\}$  that must verify  $\sum_{k=0}^{\infty} m_k^{-1} < \infty$ . Under assumptions (M1)-(M3) and these two supplementary assumptions on the MCMC procedure, and in the special case where the observed data likelihood has only one stationary point  $\hat{\theta}$ , Theorem 3 of Fort and Moulines (2003) states that the sequence of parameter updates produced by the MCMC-EM algorithm  $(\theta^k)$  converges almost surely

towards  $\hat{\theta}$ .

The Markov chains can be obtained using classical MCMC algorithms, (see, e.g., Robert and Casella (1999)). Care must also be taken regarding the specification of the Markov chain sample size. Indeed, if this size remains constant at each iteration of the algorithm, a persistent Monte Carlo error could prevent from the convergence of the algorithm. It is common practice to use a simulation schedule with a gradually increasing Monte Carlo sample size during the EM iterations, see, eg., Cappé et al. (2005) and the references therein. The details of our implementation are given in Section 5.

Note that the use of stochastic approximations instead of exact evaluation does not guarantee anymore an increase of the  $Q$ -function at each iteration. Nevertheless, solutions for this problem have already been proposed. We refer to (Caffo et al., 2005) for the general concepts regarding the ascent based Monte Carlo EM algorithm, and to Trevezas et al. (2013) for an adaptation of this algorithm in the context of plant growth models.

## 4.2 The SAEM algorithm

### 4.2.1 Description

In the MCMC-EM algorithm presented in the previous section, a new Markov Chain is generated at each iteration of the algorithm, with a size that increases with the number of iterations. Thus, at each step  $k$  of the algorithm, a new chain is built, implying that all the previous simulations are dropped. In the SAEM algorithm introduced by Delyon et al. (1999), all the previously simulated values are used for the estimation of the  $Q$ -function, even if they are gradually forgotten depending on a sequence of step sizes. The convergence of the algorithm was demonstrated in the case of i.i.d. simulations by (Delyon et al., 1999), and extended to the case of Markov Chain samples by Kuhn and Lavielle (2004).

For the non-linear mixed model described in Section 2.2, at iteration  $k$  and after the simulation of the current Markov Chain of size  $m_k$ , we update the sufficient statistics using the stochastic approximation counterparts of (34).

Each iteration  $k$  of the algorithm consists in the simulation of a Markov Chain of size  $m_k$ , with stationary distribution  $f(\phi_i | y_i)$ , and the sufficient statistics are then updated in the following way:

$$t_1^{(k)} = t_1^{(k-1)} + \gamma_k \left[ \sum_{i=1}^s \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_{i,1} \left( \phi_i^{k,(m)} \right) \right) - t_1^{(k-1)} \right] \quad (37)$$

$$t_2^{(k)} = t_2^{(k-1)} + \gamma_k \left[ \sum_{i=1}^s \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_{i,2} \left( y_i, \phi_i^{k,(m)} \right) \right) - t_2^{(k-1)} \right]. \quad (38)$$

The missing information matrix  $\mathcal{I}_m(\hat{\theta})$  is approximated by:

$$\hat{\mathcal{I}}_m^{(k)}(\theta; y) = \hat{\mathcal{I}}_m^{(k-1)}(\theta; y) + \gamma_k \left[ \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} t_m^{(k)'} - \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} \right) \left( \frac{1}{m_k} \sum_{m=1}^{m_k} t_m^{(k)} \right)' - \hat{\mathcal{I}}_m^{(k-1)}(\theta; y) \right]. \quad (39)$$

As in the MCMC-EM algorithm, the same approaches can be used to compute the Markov Chains (see Section 4.3).

#### 4.2.2 Convergence issues

As with the MCMC-EM algorithm, more assumptions are needed in addition to hypothesis (M1)-(M3) of Section 3.3 to ensure the convergence of the SAEM algorithm. These assumptions were given by Delyon et al. (1999) when the E step is explicit, and were completed by Kuhn and Lavielle (2004) when the SAEM is coupled with a MCMC procedure, in the case where the E step is not explicit.

The first hypothesis concerns the sequence of step sizes  $\{\gamma_k\}$  that should satisfy for all positive integers  $k$ ,  $\gamma_k \in [0, 1]$ ,  $\sum_{k=1}^{\infty} \gamma_k = \infty$  and  $\sum_{k=1}^{\infty} \gamma_k^{1+\lambda} < \infty$  for some  $\lambda \in (1/2, 1]$ . Delyon et al. (1999) showed that an optimal speed of convergence can be obtained with  $\gamma_k \propto k^{-a}$ ,  $1/2 < a \leq 1$ . However, if a large step size may lead to a quicker convergence, it also induces a bigger Monte Carlo error. Conversely, a small step size allows for a smaller Monte Carlo error, but can slow down the convergence. Kuhn and Lavielle (2005) suggested to start with a step of size 0, i.e.  $\gamma_k = 1$ , to ensure a quick convergence towards a neighborhood of the maximum likelihood, and to decrease the step size once we are near enough to the maximum likelihood, to ensure an almost sure convergence of the algorithm. For  $K_1$  iterations, we let  $a = 0$ , and for the next  $K_2$  iterations the step size is decreased:

$$\gamma_k = \begin{cases} 1 & \text{for } 1 \leq k \leq K_1 \\ \frac{1}{k - K_1 + 1} & \text{for } k > K_1. \end{cases} \quad (40)$$

No clear criterion exists to choose  $K_1$  and  $K_2$ . Kuhn and Lavielle (2004) suggested that in practice,  $50 < K_1 < 100$  is enough to ensure the convergence to the neighbourhood of the solution. Of course, such empirical observations strongly depend on the context, and it seems more appropriate to evaluate this parameter on the data, for example via a graphical check, or, to develop automated methods like the one proposed by Jank (2006b) for example.

The second hypothesis for the convergence of the SAEM algorithm requires that the observed-data log-likelihood  $\ell : \Theta \mapsto \mathbb{R}$  and the function  $\hat{\theta} : \mathcal{T} \mapsto \Theta$  are  $d$ -times differentiable, where  $d$  is the dimension of the sufficient statistics. In our case, as  $L$  is the integral of a product of two Gaussian densities, which are indefinitely differentiable, and as inversion of differentiation and integration is valid,  $L$  is also indefinitely differentiable and in particular,  $d$ -times differentiable. As for function  $\hat{\theta}$ , it is obvious from expression (20) that it is  $d$ -times differentiable.

The third condition concerns the transition probability of the MCMC procedure used in the stochastic approximation step. Firstly, the chain must take its values in a compact subset of  $\mathbb{R}^P$ . Then, considering the probability transition  $\pi_\theta(x, y)$  as a function of  $\theta$ , it must be a Lipschitz function, whose constant is uniform over  $(x, y)$ . Third, the transition probability must produce a uniformly ergodic chain. Finally, function  $\bar{t}$  must be bounded on the compact subset on which the chain is taking its values.

A ultimate condition is needed, and states that the sequences of sufficient statistics  $t_1^{(k)}$  and  $t_2^{(k)}$  must stay within a compact subset of  $\mathcal{T}$ . However, according to the authors, this condition may be difficult to check, or may even not be valid, in which cases it is always possible to consider a stabilisation of the algorithm by reinitializing it whenever the sequence falls outside of a given compact subset.

### 4.3 Construction of the Markov Chains

Two MCMC algorithms were used to generate samples from the distribution of the hidden data given the observed ones: the Metropolis-Hastings algorithm, and the hybrid Gibbs Sampler also known as ‘‘Metropolis-Within-Gibbs’’ (Tierney, 1994). For each of these two algorithms, two different proposals will be used, either the marginal distribution of the hidden data, or a gaussian random walk. We consider in the sequel the resulting expressions for the acceptance probabilities. It is known that whatever the choice of the proposal distribution, the Markov Chain will converge to the target distribution as long as some very general conditions are verified, nevertheless the choice of the proposal can influence the speed of convergence. In particular, several adaptive algorithms have been proposed and are supposed to provide optimal performances. We refer the reader to Andrieu and Thoms (2008) for a review of the existing approaches.

#### 4.3.1 Metropolis-Hastings

Using a proposal distribution  $q(y | x)$ , the  $m$ -th iteration of the Metropolis-Hastings algorithm consists in drawing a candidate  $\tilde{\phi}_i \sim q(\cdot | \phi_i^{k,(m)})$ , and accept it as the  $m$ -th element of the chain with probability  $\alpha(\phi_i^{k,(m)}, \tilde{\phi}_i)$ , defined as:

$$\alpha(\phi_i^{k,(m)}, \tilde{\phi}_i) = \min \left( 1, \frac{f(\tilde{\phi}_i | y_i; \theta^k) q(\phi_i^{k,(m)} | \tilde{\phi}_i)}{f(\phi_i^{k,(m)} | y_i; \theta^k) q(\tilde{\phi}_i | \phi_i^{k,(m)})} \right), \quad (41)$$

or equivalently,

$$\phi_i^{k,(m+1)} = \begin{cases} \tilde{\phi}_i & \text{with probability } \alpha(\phi_i^{k,(m)}, \tilde{\phi}_i), \\ \phi_i^{k,(m)} & \text{with probability } 1 - \alpha(\phi_i^{k,(m)}, \tilde{\phi}_i). \end{cases} \quad (42)$$

- When using the marginal distribution  $f(\phi_i; \theta^k)$  (i.e. the density of a multivariate Gaussian  $\mathcal{N}(\beta^{(k)}, \Gamma^{(k)})$ , where  $\beta^{(k)}$  and  $\Gamma^{(k)}$  are the current estimation of  $\beta$  and  $\Gamma$ ) as the proposal, the acceptance proba-

bility (41) simplifies as follows:

$$\alpha(\phi_i^{k,(m)}, \tilde{\phi}_i) = \frac{f(\tilde{\phi}_i | y_i; \theta^k) f(\phi_i^{k,(m)}; \theta^k)}{f(\phi_i^{k,(m)} | y_i; \theta^k) f(\tilde{\phi}_i; \theta^k)} = \frac{f(y_i | \tilde{\phi}_i; \theta^k)}{f(y_i | \phi_i^{k,(m)}; \theta^k)}. \quad (43)$$

- When using a symmetric random walk, for example a Gaussian random walk, the ratio (41) simplifies as follows (see Andrieu and Thoms (2008) or Haario et al. (2001)):

$$\alpha(\phi_i^{k,(m)}, \tilde{\phi}_i) = \frac{f(\tilde{\phi}_i | y_i; \theta^k)}{f(\phi_i^{k,(m)} | y_i; \theta^k)} = \frac{f(y_i | \tilde{\phi}_i; \theta^k) f(\tilde{\phi}_i; \theta^k)}{f(y_i | \phi_i^{k,(m)}; \theta^k) f(\phi_i^{k,(m)}; \theta^k)}. \quad (44)$$

### 4.3.2 Hybrid Gibbs sampler

In the Gibbs sampler, each component of  $\phi_{i,j}$  is simulated using the corresponding full conditional distribution  $f_j(\phi_{i,j} | \phi_{i,1}, \dots, \phi_{i,j-1}, \phi_{i,j+1}, \dots, \phi_{i,P}, y_i; \theta)$ . However, simulation from the full conditional distributions may not be straightforward. The hybrid Gibbs sampler, also called ‘‘Metropolis-within-Gibbs’’, consists in substituting simulations from the full conditional distributions of a Gibbs Sampler, by simulations from an instrumental distribution. Given  $(\phi_{i,1:j-1}^{k,(m+1)}, \phi_{i,j:P}^{k,(m)})$ , where  $\phi_{i,p:q}^{k,(m)}$  is a shorthand for  $(\phi_{i,p}^{k,(m)}, \dots, \phi_{i,q}^{k,(m)})$ , the  $m + 1$ -th step of the algorithm is defined as:

1. Simulate  $\tilde{\phi}_{i,j} \sim q_j(\cdot | \phi_{i,1:j-1}^{k,(m+1)}, \phi_{i,j:P}^{k,(m)}; y_i; \theta)$
2. Take

$$\phi_{i,j}^{k,(m+1)} = \begin{cases} \phi_{i,j}^{k,(m)} & \text{with probability } 1 - \alpha_j, \\ \tilde{\phi}_{i,j} & \text{with probability } \alpha_j, \end{cases} \quad (45)$$

where

$$\alpha_j = \min \left\{ 1, \frac{f_j(\tilde{\phi}_{i,j} | \phi_{i,1:j-1}^{k,(m+1)}, \phi_{i,j+1:P}^{k,(m)}; y_i; \theta) q_j(\phi_{i,j}^{k,(m)} | \phi_{i,1:j-1}^{k,(m+1)}, \tilde{\phi}_{i,j}, \phi_{i,j+1:P}^{k,(m)}; y_i; \theta)}{f_j(\phi_{i,j}^{k,(m)} | \phi_{i,1:j-1}^{k,(m+1)}, \phi_{i,j+1:P}^{k,(m)}; y_i; \theta) q_j(\tilde{\phi}_{i,j} | \phi_{i,1:j-1}^{k,(m+1)}, \phi_{i,j}^{k,(m)}, \phi_{i,j+1:P}^{k,(m)}; y_i; \theta)} \right\}. \quad (46)$$

As in the case of the Metropolis-Hastings algorithm, we compared two types of proposals, and give the simplifications of the acceptance probabilities in both cases:

- When using the marginal distribution  $f(\phi_{i,j}; \theta^k)$  the acceptance probability (46) simplifies as follows:

$$\alpha_j = \frac{f(y_i | \tilde{\phi}_{i,j}; \theta^k)}{f(y_i | \phi_{i,j}^{k,(m)}; \theta^k)} \quad (47)$$

- When using a symmetric random walk, for example a Gaussian random walk, the ratio (46) simplifies as follows:

$$\alpha_j = \frac{f(y_i | \tilde{\phi}_{i,j}; \theta^k) f(\tilde{\phi}_{i,j}; \theta^k)}{f(y_i | \phi_{i,j}^{k,(m)}; \theta^k) f(\phi_{i,j}^{k,(m)}; \theta^k)}. \quad (48)$$

## 5 Results

In this section we apply the two algorithms described previously on simulated and real data sets from the sugar beet.

### 5.1 Simulated data

A global sensitivity analysis of Sobol type (Wu et al., 2012) was applied to the Greenlab model in (Baey et al., 2013). The results revealed that the three most influential parameters are  $\mu$ ,  $s^{pr}$  and  $a_r$ , with high total-order indices. Due to the complexity of our model, we first only consider these three parameters as random effects. The other parameters are 'screened' (Campolongo et al., 2007) and fixed to their mean values in the uncertainty interval taken for the sensitivity analysis. Therefore, a first set of 50 vectors of individual parameters were generated, that lead to the simulation of 50 virtual plants. As the three parameters considered are by assumption strictly positive, we used a logarithmic transformation as stated in Section 2.2, i.e. we assumed log-normal distributions for these three random parameters. In the sequel, we denote  $\beta_0 = \mathbb{E}(\log \mu)$ ,  $\beta_1 = \mathbb{E}(\log s^{pr})$ ,  $\beta_2 = \mathbb{E}(\log a_r)$ ,  $\sigma_0 = \text{sd}(\log \mu)$ ,  $\sigma_1 = \text{sd}(\log s^{pr})$ , and  $\sigma_2 = \text{sd}(\log a_r)$ .

Ten independent realizations of MCMC-EM and SAEM were then launched for each combination of the MCMC algorithm (Metropolis-Hastings –MH– or Hybrid Gibbs Sampler –hGs–) and the proposal (marginal or adaptive random walk). For each independent run, initial values for the algorithm were sampled from a uniform distribution with a range of  $\pm 20\%$  around the true value of the corresponding parameter. This allowed us to have a first estimate of the Monte Carlo error, but it should be noted that in practice, only one realization of the algorithm is necessary.

Results among all the algorithms were consistent, but for the sake of clarity, we present here the results from the MH algorithm with a symmetric adaptive random walk, as it was the more efficient combination. In Table 1, we give the results for the parameter estimation of  $\theta_1$  and  $\theta_2$ , averaged from the ten independent realizations of both algorithms. More specifically: in the second column we give the true values of the parameters that were used to simulate the data ; in the third column we give the estimation of the parameters based on the fully observed data, i.e. using the individual parameters generated to simulate the 50 plants as if they were observed ; in columns 4 and 6 we provide the averaged estimates of the parameters obtained by the SAEM and the MCMC-EM algorithm respectively ; and in columns 5 and 7 we give the averaged standard errors computed by the Louis's missing principle, obtained by the SAEM and the MCMC-EM algorithm respectively.

Optimal acceptance rates, between 0.3 and 0.5 and with a mean value of 0.41 at the last iteration, were obtained when a global adaptive scaling was used, as proposed in Andrieu and Thoms (2008). This is higher than the theoretically optimal acceptance rate for multivariate settings, which has been shown

to be equal to 0.234 (Roberts et al., 1997), but it is worth noting that this optimal rate is valid only asymptotically, when the dimension of the Markov chain tends to infinity, and for i.i.d. components, and may not apply in our case.

Table 1: Results from the SAEM and the MCMC-EM algorithms on the simulated dataset. The values used to generate the data are given in the first column, and estimates from the fully observed data are given in the second column. Then, for each algorithm, we give the estimates and standard errors computed by Louis’s missing information principle, averaged from the ten independent realizations.

Parameter	True value	Fully observed	SAEM		MCMC-EM	
			Estimate	Standard error	Estimate	Standard error
$\beta_0$	1.7	1.7036	1.6976	0.0210	1.6981	0.0210
$\sigma_0$	0.15	0.1544	0.1484	0.0148	0.1484	0.0148
$\beta_1$	-3	-2.9062	-2.9181	0.0823	-2.9159	0.0820
$\sigma_1$	0.5	0.5729	0.5818	0.0582	0.5787	0.0586
$\beta_2$	1.45	1.4906	1.5002	0.0162	1.4992	0.0163
$\sigma_2$	0.15	0.1232	0.1149	0.0115	0.1154	0.0115
$\sigma_b^2$	0.15	0.2019	0.2033	0.0055	0.2035	0.0055
$\sigma_r^2$	0.15	0.0827	0.0842	0.0023	0.0844	0.0023
$\rho$	0.67	0.5342	0.5290	0.0139	0.5292	0.0139
$\sigma_r^2$	0.15	0.1389	0.1231	0.0246	0.1276	0.0255

As a simulation schedule we opted for the one used in the Monte Carlo versions of the ECM algorithm presented in Trevezas and Cournède (2013) and Trevezas et al. (2013), which concern parameter estimation in the case of the individual-based Greenlab model. In particular, after a first period of piecewise linear increase, the parameters are driven towards their optimum values by increasing the chain size quadratically with the EM iteration number (see also Cappé et al. (2005)). In order to smooth the parameter estimates during the final iterations of the EM algorithm (near the convergence region) we used an averaging technique, as explained for example in Fort and Moulines (2003). We are currently working on an automated version of this algorithm in order to extend the use of the automated algorithm developed in Trevezas et al. (2013) for the individual-based Greenlab model to the population Greenlab one introduced here. For the SAEM algorithm, since the simulation schedule is determined by the choice of two parameters as described in Section 4.2.2,  $K_1$  was set to 100 and  $K_2$  to 70.

Results from the two algorithms are shown in Fig. 2 and 3 respectively. We do not show the results for  $\theta_2$  as convergence is reached very quickly for these parameters due to the higher number of observations available for their estimation (except for  $\sigma_r^2$ ). Globally, the mean parameters  $\beta$  are better estimated than the variance parameters  $\sigma$ , with a smaller variability. Mean computation time was 10 times smaller with the SAEM algorithm. Not however that we did not use an automated version of the MCMC-EM, which could have saved us some time since the resources are supposed to be used in a more efficient way with the automated version. We can see that both algorithms give very similar values for the estimates and their standard errors, but we observe a higher variability between the independent runs of the SAEM algorithms, especially for the variance components.

Table 2 gives the comparison between the standard errors of the estimates computed by the Louis’s



Table 2: Comparison of the standard errors of the estimates, computed either by parametric bootstrap (100 samples, column 3) or by Louis's missing principle (column 6), using the SAEM algorithm

Parameter	Bootstrap			FIM		
	Estimate	SD	IC	Estimate	SD	IC
$\beta_0$	1.7001	0.0191	[1.664 ; 1.737]	1.6976	0.0210	[1.657 ; 1.739]
$\sigma_0$	0.1488	0.0239	[ 0.108 ; 0.179]	0.1484	0.0148	[0.119 ; 0.178]
$\beta_1$	-2.9182	0.0891	[-3.096 ; -2.705]	-2.9181	0.0823	[-3.079 ; -2.757]
$\sigma_1$	0.5930	0.0670	[0.450 ; 0.706]	0.5818	0.0582	[0.468 ; 0.696]
$\beta_2$	1.5039	0.0291	[1.445 ; 1.564]	1.5003	0.0162	[1.468 ; 1.532]
$\sigma_2$	0.2004	0.0197	[0.151 ; 0.235]	0.1149	0.0115	[0.092 ; 0.137]
$\sigma_b^2$	0.2285	0.0057	[0.215 ; 0.241]	0.2033	0.0055	[0.192 ; 0.214]
$\sigma_p^2$	0.0381	0.0022	[0.034 ; 0.042]	0.0842	0.0023	[0.080 ; 0.089]
$\rho$	0.6542	0.0219	[0.598 ; 0.696]	0.5290	0.0139	[0.502 ; 0.556]
$\sigma_r$	0.1604	0.1089	[-0.160 ; 0.252]	0.1231	0.0246	[0.075 ; 0.171]

missing principle, as detailed in Section 3.4, and by the parametric bootstrap. Due to the similar if not identical results obtained with the MCMC-EM and the SAEM on these standard errors, and given the much smaller computation time of the latter, we did the comparison using the SAEM algorithm. The bootstrap sample size was set to 100, which is lower than what is generally used in these contexts, but the computation cost, although cheaper with the SAEM, remains high. If the results are quite consistent for the mean parameters, except for  $\beta_2$ , the variability of the SAEM algorithm for the variance components of the random effects lead to higher standard errors computed with the parametric bootstrap, compared to those calculated with the Fisher Information Matrix. It is also particularly true for the variance parameter  $\sigma_r^2$ . We can also observe some discrepancies between the estimates of the observation noises, especially for the covariance parameter  $\rho$ . This higher variability could certainly be reduced with a larger bootstrap size. Nevertheless, these results suggest that the method based on the Fisher Information Matrix is very promising and gives satisfactory results.

## 5.2 Real data

Real data comes from 2010 experiments conducted by the French institute for sugar beet research (ITB, Paris) at La Selve in France on 18 plants. Dry matter of each blade and each petiole for each individual plant, as well as the root biomass were measured destructively at day 160. The experimental protocol is detailed in Baey et al. (2013).

In practice, a single run from each candidate stochastic algorithm is enough to estimate the parameters of the model with satisfying precision. Indeed, the Monte Carlo error, which can be assessed by the variability between independent runs, was shown to be small on the simulated dataset. However, in order to ensure the consistency of the results concerning the real dataset, without increasing significantly the computational cost, three independent runs from each algorithm were launched. Results from the two algorithms on real data are shown in Table 3. As suggested by the satisfactory results obtained with the Fisher information matrix on the simulated dataset, this method was used here for the computation

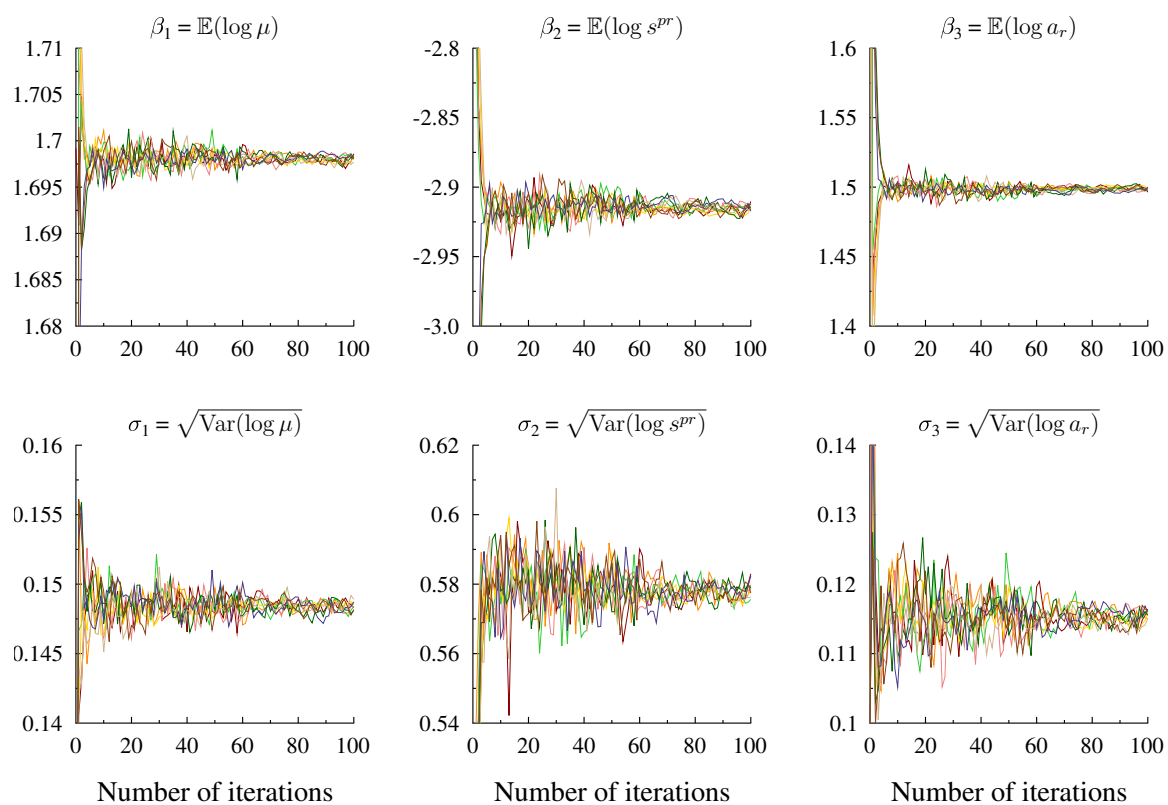


Figure 2: Results for  $\theta_1$  from ten independent realizations of the MCMC-EM algorithm on simulated data with a quadratic increase of the chain size at each iteration, and an averaging procedure from iteration 60.

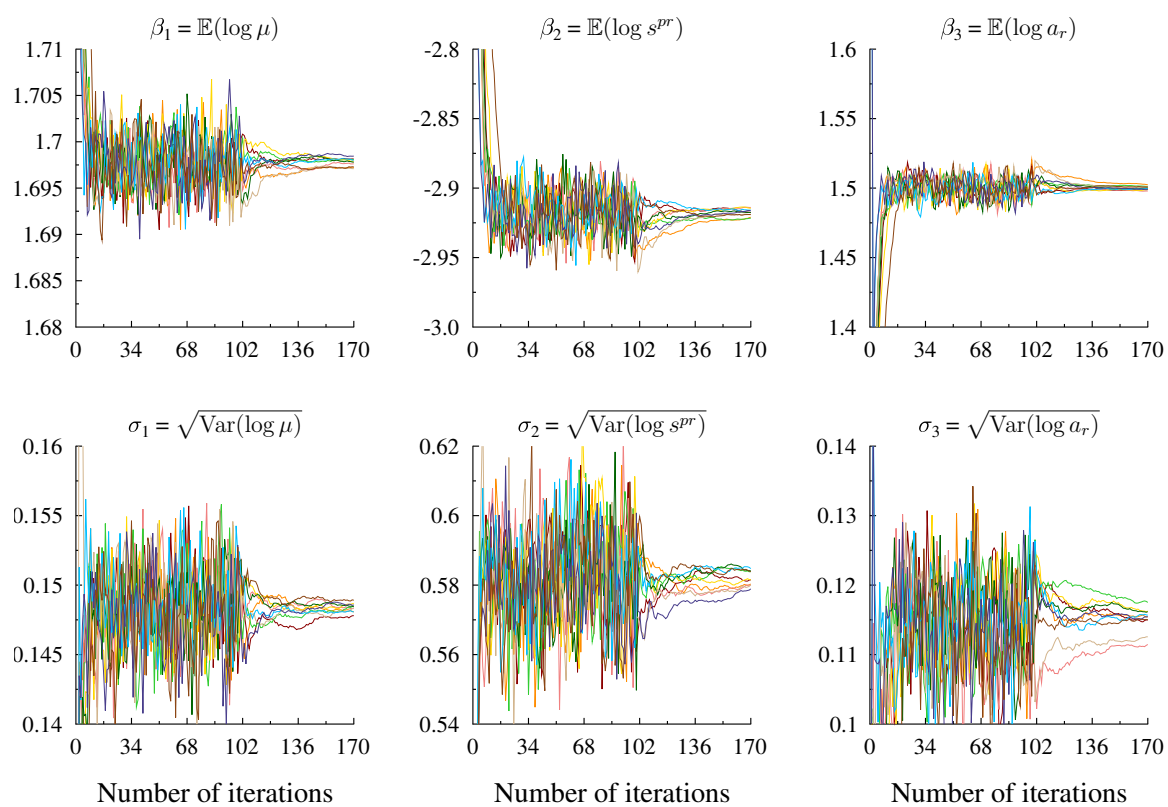


Figure 3: Result for  $\theta_1$  from ten independent realizations of the SAEM algorithm on simulated data, with  $K_1 = 100$  and  $K_2 = 70$ .

of the standard error of the estimates. The same parameters as with the simulated data were considered, i.e.  $\mu$ ,  $s^{pr}$  and  $a_r$ . The results are satisfactory, even if the standard errors are bigger than for the virtual case, due to the smaller sample size (18 in the real case versus 50 in the virtual case) and perhaps to a bigger variability than in the simulated data set. Results from the MCMC-EM showed better stability than with the SAEM algorithm, since one run of the latter algorithm provided outlying results. This run was discarded and replaced with a new one. We remark that the results are quite consistent between the two algorithms, except for the variance component of the third parameter,  $\log a_r$ . Since the variance is very close to zero, it may suggest that this parameter should not be treated as a random parameter, but rather as a fixed effect.

Table 3: Results from the SAEM and the MCMC-EM algorithms on the real dataset. Columns 2 to 5 give the average over the three independent realizations of the algorithms, of the estimates and standard errors for each parameter.

Parameter	SAEM		MCMC-EM	
	Estimate	Standard error	Estimate	Standard error
$\beta_0$	2.8188	0.0718	2.8848	0.0713
$\sigma_0$	0.3044	0.0507	0.3027	0.0505
$\beta_1$	-3.2040	0.1388	-3.2510	0.1386
$\sigma_1$	0.5885	0.0984	0.5842	0.1026
$\beta_2$	0.6461	1.07E-04	0.6203	0.0036
$\sigma_2$	0.00045	7.53E-05	0.0152	0.0025
$\sigma_b^2$	1.2078	0.0710	1.2139	0.0829
$\sigma_b^2$	1.3729	0.0808	1.3776	0.0943
$\rho$	0.97630	0.0019	0.9765	0.0021
$\sigma_r^2$	3.0096	1.1553	3.0365	1.1613

Figure 4 for MCMC-EM and Figure 5 for SAEM give the values of  $\theta_1$  according to the number of iterations. We can clearly see that for the SAEM algorithm, it may happen that a given realization does not converge as quickly as expected, for given values of  $K_1$  and  $K_2$ . It is therefore recommended to launch the algorithm several times, and maybe to discard any outlying run, or try to use an automated version of the algorithm.

## 6 Conclusion

In this study, we showed how an individual-based functional-structural plant growth model could be extended at population level to take into account inter-individual variability. The individual-based model is cast in the framework of hierarchical mixed-effects model, to represent the variability of parameters in the population.

We also derived stochastic variants of an EM-type algorithm (Expectation-Maximization) to perform maximum likelihood estimation for this type of models. The complete data distribution was shown to belong to a subclass of the exponential family of distributions for which the M-step can be solved explicitly.

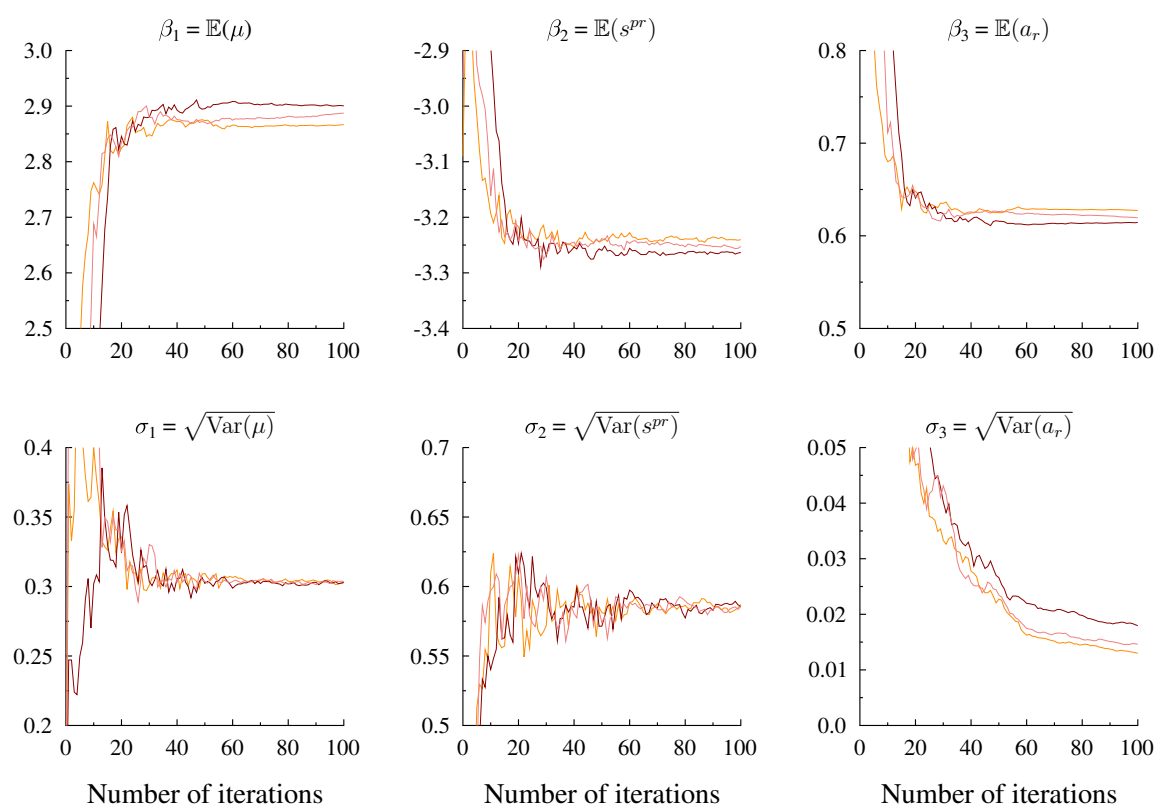


Figure 4: Results from the MCMC-EM algorithm on the real dataset with a quadratic increase of the chain size at each iteration, and an averaging procedure from iteration 60.

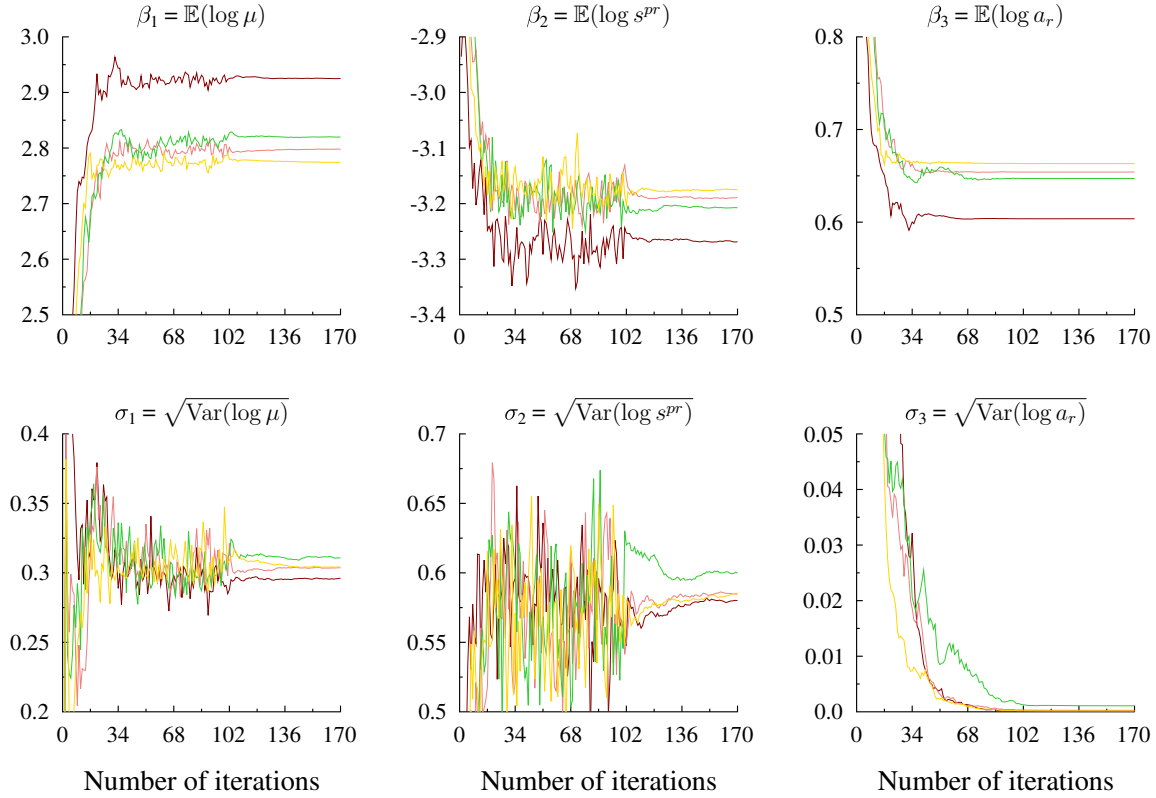


Figure 5: Results from independent realizations of the SAEM algorithm on the real dataset.

For virtual data and real experimental data, we compared two commonly used stochastic algorithms relying on Markov chain Monte-Carlo methods: the Monte-Carlo EM (MCEM) and the SAEM algorithm. For both, different versions of MCMC algorithms (Metropolis-Hastings or Hybrid Gibbs Sampler) and proposal distributions (marginal or adaptive random walk) were also tested, with consistent results. In our tests, the Metropolis-Hastings with adaptive random walk performed better.

Both MCEM and SAEM performed well, and provided satisfactory results. In our tests, SAEM was more computationally efficient (about ten times). Nevertheless, a fully automated version of the MCEM algorithm (as for example described in Trevezas et al. (2013)) could be more computationally efficient than the current one which uses a deterministic schedule in the augmentation of the Monte Carlo sample size. The choice of the proposal distribution could also have an impact on the efficiency of the algorithm, and several adaptive schemes could be used and compared to identify the most appropriate one. Since the computational burden of a population-based model is significantly larger than an individual-based one, we believe that more research effort is needed in the direction of optimizing Monte Carlo resources.

In the real data situation, MCEM was also more stable. Even if it was not the initial purpose of our paper, results on the real data set clearly showed the need to account for parameters that should only be considered as fixed effects. This should be taken into account, by integrating quasi-Newton methods to estimate the corresponding effects. Moreover, likelihood ratio tests can be performed to test whether

a parameter should be treated as random or should be considered as a fixed effect. Estimates of the log-likelihood can be derived using importance sampling methods. This will be further investigated when real data sets from larger plant populations will be available.

Concerning the SAEM, it seems very important to implement an automated version of this algorithm, since a deterministic determination of  $K_1$  and  $K_2$  may not be appropriate in some cases. This algorithm seems to be very promising in our case, since the computational cost can increase dramatically with the complexity of the model and with the dimension of the estimated parameter vector.

Finally, in the very active research field of plant growth modelling, where the ambition is to go further and further in the detailed description of physiological processes, in order to be able to link plant genetics and parameters of integrative plant models (Yin and Struik, 2010), the approach which consists in considering an average plant without accounting for the inter-individual variability is clearly bound to fail. This paper introduces for the first time a population functional-structural model, able to capture the behaviour of the population (typically for a given variety or cultivar) while taking into account the inter-individual variability. It presents the statistical framework, proposes a proper parameter estimation method, with its implementation and tests on synthetic data, and makes the first step towards validation with a real data set. However, this validation step has to be further developed. The model and methods are now tested for other species (particularly with the GreenLab model for winter oilseed rape Jullien et al. (2011)), with richer plant material and a more elaborate ecophysiological model.

## References

- Andrieu, C. and Moulines, É. (2006). On the ergodicity properties of some adaptive MCMC algorithms. *The Annals of Applied Probability*, 16(3):1462–1505.
- Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, 18:343–373.
- Baey, C., Didier, A., Lemaire, S., Maupas, F., and Cournède, P.-H. (2013). Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model. *Ecological Modelling*, 263:56–63.
- Brouwer, J., Fussell, L. K., and Herrmann, L. (1993). Soil and crop growth micro-variability in the west african semi-arid tropics: a possible risk-reducing factor for subsistence farmers. *Agriculture, Ecosystems and Environment*, 45(3-4):229 – 238.
- Caffo, B. S., Jank, W., and Jones, G. L. (2005). Ascent-based Monte Carlo Expectation-Maximization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):235–251.
- Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling and Software*, 22:1509–1518.

- Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer.
- Chen, Y., Trevezas, S., and Cournède, P.-H. (2013). A regularized particle filter em algorithm based on gaussian randomization with an application to plant growth modeling. *Submitted*.
- Cournède, P.-H., Kang, M.-Z., Mathieu, A., Barczi, J.-F., Yan, H.-P., Hu, B.-G., and de Reffye, P. (2006). Structural factorization of plants to compute their functional and architectural growth. *Simulation*, 82(7):427–438.
- Cournède, P.-H., Letort, V., Mathieu, A., Kang, M.-Z., Lemaire, S., Trevezas, S., Houllier, F., and de Reffye, P. (2011). Some Parameter Estimation Issues in Functional-Structural Plant Modelling. *Mathematical Modelling of Natural Phenomena*, 6(2):133–159.
- Cournède, P.-H., Mathieu, A., Houllier, F., Barthélémy, D., and de Reffye, P. (2008). Computing competition for light in the Greenlab model of plant growth: a contribution to the study of the effects of density on resource acquisition and architectural development. *Annals of Botany*, 101(8):1207–1219.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall/CRC, 1st edition.
- Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4):387–419.
- de Reffye, P. and Hu, B.-G. (2003). Relevant qualitative and quantitative choices for building an efficient dynamic plant growth model: GreenLab case. In Hu, B. G. and Jaeger, M., editors, *First International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA)*, Beijing, China, pages 87–107. Tsinghua University Press and Springer.
- Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27(1):94–128.
- Dempster, A., Laird, N. M., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.
- Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus Expected Fisher information. *Biometrika*, 65(3):457 – 487.
- Fort, G. and Moulines, E. (2003). Convergence of the monte carlo expectation maximization for curved exponential families. *The Annals of Statistics*, 31(4):1220–1259.



- Fournier, C. and Andrieu, B. (1999). ADEL-maize: An L-system based model for the integration of growth processes from the organ to the canopy. Application to regulation of morphogenesis by light availability. *Agronomy*, 19(3-4):313–327.
- Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242.
- Jank, W. (2005). Stochastic Variants of EM: Monte Carlo, Quasi-Monte Carlo and More. In *Proceedings of the American Statistical Association*.
- Jank, W. (2006a). The EM algorithm, Its Stochastic Implementation and Global Optimization: Some Challenges and Opportunities for OR. In Alt, F., Fu, M., and Golden, B., editors, *Topics in Modeling, Optimization and Decision Technologies: Honoring Saul Gass' Contributions to Operation Research*, pages 367–392. Springer-Verlag.
- Jank, W. (2006b). Implementing and diagnosing the stochastic approximation EM algorithm. *Journal of Computational and Graphical Statistics*, 1815:1–30.
- Jullien, A., Mathieu, A., Allirand, J.-M., Pinet, A., de Reffye, P., Cournède, P.-H., and Ney, B. (2011). Characterisation of the interactions between architecture and source:sink relationships in winter oilseed rape (*brassica napus* l.) using the greenlab model. *Annals of Botany*, 107(5):765–779.
- Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics*, 8:115–131.
- Kuhn, E. and Lavielle, M. (2005). Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 49(4):1020–1038.
- Lemaire, S., Maupas, F., Cournède, P.-H., and de Reffye, P. (2008). A morphogenetic crop model for sugar-beet (*beta vulgaris* l.). In *International Symposium on Crop Modeling and Decision Support: ISCMDS 2008, April 19-22, 2008, Nanjing, China*.
- Louis, T. A. (1982). Finding the Observed Information Matrix when using the EM-algorithm. *Journal of the Royal Statistical Society*, 44(2):226–233.
- McCulloch, C. E. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89(425):330–335.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American statistical Association*, 92(437):162–170.
- McLachlan, G. and Krishnan, T. (2007). *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. Wiley, 2nd edition.

- Orchard, T. and Woodbury, M. (1972). A missing information principle: theory and applications. In Le Cam, L., Neyman, J., and Scott, J., editors, *Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1: Theory of Statistics. University of California Press.
- Racine-Poon, A. (1985). A bayesian approach to nonlinear random effects models. *Biometrics*, pages 1015–1023.
- Renno, J.-F. and Winkel, T. (1996). Phenology and reproductive effort of cultivated and wild forms of *Pennisetum glaucum* under experimental conditions in the Sahel : implications for the maintenance of polymorphism in the species. *Canadian Journal of Botany*, 74:959–964.
- Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.*, 22:400–407.
- Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Texts in Statistics Series. Springer-Verlag GmbH.
- Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120.
- Sievänen, R., Nikinmaa, E., Nygren, P., Ozier-Lafontaine, H., Perttunen, J., and Hakula, H. (2000). Components of functional-structural tree models. *Annals of Forest Science*, 57(5):399–412.
- Sundberg, R. (1974). Maximum likelihood theory for incomplete data from an exponential family. *Scandinavian Journal of Statistics*, 1:49–58.
- Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, pages 1701–1728.
- Trevezas, S. and Cournède, P. (2013). A sequential Monte Carlo approach for MLE in a plant growth model. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2):250–270.
- Trevezas, S., Malefaki, S., and Cournède, P.-H. (2013). Simulation techniques for parameter estimation via a stochastic ECM algorithm with applications to plant growth modeling. *Submitted*.
- Vos, J., Marcelis, L., and Evers, J. (2007). *Functional-structural plant modelling in crop production*, chapter 1. Springer.
- Wei, G. C. G. and Tanner, M. A. (1990). A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85(411):699–704.
- Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103.

Wu, Q., Cournède, P.-H., and Mathieu, A. (2012). An efficient computational method for global sensitivity analysis and its application to tree growth modelling. *Reliability Engineering & System Safety*, 107:35–43.

Yin, X. and Struik, P. C. (2010). Modelling the crop: from system dynamics to systems biology. *Journal of Experimental Botany*, 61(8):2171–2183.

## Appendix

### Proof of Proposition 2.

Let us first define:

$$\Sigma(\beta) := \left( \sum_{i=1}^s n_i \right)^{-1} \sum_{i=1}^s \sum_{n=1}^{n_i} (y_{i,n} - \log G_n(\beta)) (y_{i,n} - \log G_n(\beta))^t, \quad (49)$$

$$\sigma_r^2(\beta) := \frac{1}{s} \sum_{i=1}^s (y_{i,1}^r - \log G_1^r(\beta))^2. \quad (50)$$

Now, we give the following representation of the observed-data likelihood:

$$L(\theta) = \int_{\mathbb{R}^{P \times s}} f(\phi; \theta_1) f(y | \phi; \theta_2) d\phi = \mathbb{E}_{\theta_1} [f(y | \Phi; \theta_2)], \quad (51)$$

where  $\theta_1 = (\beta, \sigma^2)$ ,  $\theta_2 = (\sigma_b^2, \sigma_p^2, \sigma_r^2, \rho)$  and  $\Phi \sim f(\phi; \theta_1)$ . Let  $\theta_n = (\theta_{1,n}, \theta_{2,n})$  be a sequence that converges to  $\theta = (\theta_1, \theta_2)$ , and let us define  $h_n(\phi) := f(y | \phi; \theta_{2,n})$ ,  $h(\phi) := f(y | \phi; \theta_2)$ ,  $\Phi_n \sim f(\phi; \theta_{1,n})$  and  $\Phi \sim f(\phi; \theta_1)$ . Then, we have:

$$L(\theta_n) = \mathbb{E}_{\theta_{1,n}} [f(y | \Phi; \theta_{2,n})] = \mathbb{E} [f(y | \Phi_n; \theta_{2,n})] = \mathbb{E} [h_n(\Phi_n)], \quad (52)$$

and we will show that  $L(\theta_n) \rightarrow L(\theta)$ . We remark that the conditions of the generalized continuous-mapping theorem are verified: (i)  $h_n$  and  $h$  are measurable functions taking values in  $(\mathbb{R}, |\cdot|)$  which is a separable metric space, (ii) by continuity of  $f(y | \phi; \theta_2)$  in  $(\phi, \theta_2)$ , we have that for any  $\phi$  and any sequence  $\{\phi_n\}$  converging to  $\phi$ ,  $h_n(\phi_n) = f(y | \phi_n; \theta_{2,n}) \rightarrow f(y | \phi; \theta_2) = h(\phi)$ , (iii) since  $f(\phi; \theta_1)$  is a continuous function of  $\theta_1$ , we have the following convergence of densities  $f(\phi; \theta_{1,n}) \rightarrow f(\phi; \theta_1)$  and thus, by Scheffe's Lemma,  $\Phi_n \xrightarrow{d} \Phi \sim f(\phi; \theta_1)$ . By the generalized continuous-mapping theorem, we thus have that  $h_n(\Phi_n) \xrightarrow{d} h(\Phi)$ . In order to deduce the convergence of their expectations it is enough to show that  $\{h_n(\Phi_n)\}$  is uniformly bounded, and consequently uniformly integrable. Indeed, notice that

$$h_n(\phi) \leq \frac{(2\pi)^{-\sum_{i=1}^s n_i - s/2}}{(\det \Sigma_{b,p;n})^{\sum_{i=1}^s n_i/2} (\sigma_{r;n}^2)^{s/2}}, \quad n \geq 1, \quad (53)$$

and since  $\theta_{2,n} \rightarrow \theta_2$ , the sequence of bounds of  $h_n$  given by (53) also converges, and consequently is bounded by some  $M > 0$ . We deduce that  $\{h_n(\Phi_n)\}$  is uniformly bounded by  $M$ , and by (52) and the previous arguments, we have  $L(\theta_n) = \mathbb{E}[h_n(\Phi_n)] \rightarrow \mathbb{E}[h(\Phi)] = L(\theta)$ .

Let us consider the non linear model given by (4) but without random effects. We will show that its likelihood function  $L(\mu)$  is bounded provided that the condition stated in the Proposition is valid. The corresponding equation is given here by:

$$y_{i,n} = \log G_n(\beta) + \epsilon_{i,n}, \quad \epsilon_{i,n} \sim \mathcal{N}_{d_n}(0, \Sigma_n), \quad (54)$$

where  $\{\epsilon_{i,n}\}$  are independent as in the model with the random effects and  $\Sigma_n$  is given by (6). This model is now Gaussian and by setting  $\mu := (\beta, \theta_2)$  its likelihood function is given by

$$\begin{aligned} L(\mu) &= \frac{(2\pi)^{-\sum_{i=1}^s n_i - s/2}}{(\det \Sigma_{b,p})^{\sum_{i=1}^s n_i/2} (\sigma_r^2)^{s/2}} \times \exp \left\{ -\frac{1}{2} \sum_{i,n} (y_{i,n} - \log G_n(\beta))^t \Sigma_{b,p}^{-1} (y_{i,n} - \log G_n(\beta)) \right\} \\ &\times \exp \left\{ -\frac{1}{2\sigma_r^2} \sum_{i=1}^s (y_{i,1}^r - \log G_1^r(\beta))^2 \right\}, \end{aligned} \quad (55)$$

and will not belong to the exponential family in general, due to the non linear dependence of  $G_n$  on the parameter  $\beta$ . Nevertheless,  $\log G_n(\beta)$  is continuous in  $\beta$  and since  $L(\mu)$  is also continuous in  $\theta_2$ , we deduce by (55) that  $L$  is continuous in  $\mu$ . For a fixed  $\beta$ ,  $L(\mu)$  has a unique maximizer  $\theta_2(\beta) := (\sigma_b^2(\beta), \sigma_p^2(\beta), \rho(\beta), \sigma_r^2(\beta))$ , where  $\Sigma(\beta)$  and  $\sigma_r^2(\beta)$  are given by (49) and (50) respectively. In particular, we obtain easily that for some constant  $c > 0$  we have:

$$L(\mu) \leq \sup_{\theta_2} L(\beta, \theta_2) = \frac{c}{(\det \Sigma(\beta))^{\sum_{i=1}^s n_i/2} (\sigma_r^2(\beta))^{s/2}}. \quad (56)$$

Note that the denominator in (56) is strictly positive by the condition. Then, we conclude that

$$\sup_{\mu} L(\mu) = \sup_{\beta, \theta_2} L(\beta, \theta_2) \leq \frac{c}{\inf_{\phi} (\det \Sigma(\beta))^{\sum_{i=1}^s n_i/2} (\sigma_r^2(\beta))^{s/2}} =: M < \infty. \quad (57)$$

and  $L(\mu)$  is also bounded. The next step is to extend these results to  $L(\theta)$ . Indeed, we can show inductively that as random parameters are added to the initial model with likelihood  $L(\mu)$ , the extended model has a continuous likelihood function (the arguments are the same as in the proof of continuity of  $L(\theta)$ ). Also we remark that as  $\sigma_j^2$  converges to 0, for some random parameter  $\Phi_j \sim \mathcal{N}(\beta_j, \sigma_j^2)$ , then  $L(\theta_{-j}, \beta_j, \sigma_j^2)$  converges to  $L(\theta_{-j}, \beta_j) := \mathbb{E}_{\theta_{1,-j}} [f(y | \Phi_{-j}; \theta_2, \beta_j)]$ , where  $\theta_{-j} = \theta \setminus (\beta_j, \sigma_j^2)$ ,  $\theta_{1,-j} = \theta_1 \setminus (\beta_j, \sigma_j^2)$  and  $\Phi_{-j} = \Phi \setminus \Phi_j$ , where  $\Phi_j = (\Phi_{1j}, \dots, \Phi_{sj})$ . The proof is similar to the one given in Proposition 1 of Chen et al. (2013) and will be omitted. We deduce that the likelihood of an extended model is bounded as the variance of the corresponding random effect approaches the boundary and since

this is also true for the rest of its values, we should obtain that  $L(\theta)$  is bounded.