HAL
open science

A Regularized Particle Filter EM Algorithm Based on Gaussian Randomization with an Application to Plant Growth Modeling

Yuting Chen, Samis Trevezas, Paul-Henry Cournède

To cite this version:

Yuting Chen, Samis Trevezas, Paul-Henry Cournède. A Regularized Particle Filter EM Algorithm Based on Gaussian Randomization with an Application to Plant Growth Modeling. Methodology and Computing in Applied Probability, 2015, 17 (4), pp.847-870. 10.1007/s11009-015-9440-0 . hal-00997732

HAL Id: hal-00997732

https://hal.science/hal-00997732

Submitted on 30 May 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 regularized particle filter EM algorithm based on Gaussian randomization with an application to plant growth modeling

Yuting Chen[1], Samis Trevezas[1,*], Paul-Henry Cournède[1]

[1]Ecole Centrale Paris, Laboratoire de Mathématiques Appliquées aux Systèmes

Abstract

Parameter estimation in complex models arising in real data applications is a topic which still attracts a lot of interest. In this article, we study a specific data and parameter augmentation method which gives us the opportunity to estimate more easily the parameters of the initial model. For this reason, the notion of Gaussian randomization of a model with respect to some of its parameters is introduced. The initial model can be regarded as a submodel of the resulting extended incomplete data model. Under the assumption that the initial model has a unique maximum likelihood estimator (MLE) and that the likelihood function is continuous we prove that the extended model has a unique MLE with common values for the parameters of the MLE which correspond to the initial model. We also prove the reverse direction. Moreover, an appropriate stochastic version of an EM (Expectation-Maximization) algorithm is suggested to make parameter estimation feasible. In particular, we describe how the regularized particle filter of [21] can be used in this frequentist-based approach to perform the Monte Carlo E-step at each iteration of the stochastic EM algorithm. This regularized version is particularly adapted to the framework of Gaussian randomization since the last iterations of the EM algorithm are characterized by low variance in the parameter distributions. A toy example with available analytic solutions, a synthetic example and a real data application with scarce observations to the LNAS (Log-Normal Allocation and Senescence) model of sugar beet growth are presented to highlight some theoretical and practical aspects of the proposed methodology.

Key words and phrases: Gaussian randomization; stochastic EM algorithm; regularized particle filter; plant growth model; LNAS model; state space model

1 Introduction

The current paper is motivated by the need to have a simpler method to perform parameter estimation in plant growth models which are usually characterized by complex

*Corresponding author; e–mail: samis.trevezas@ecp.fr

parameterization and scarce experimental data. Nevertheless, our proposed method is generic and could be used in any statistical model under certain conditions that we state in the sequel. The basic idea consists in interpreting some of the parameters of the model as the means of unobserved states following normal distributions with unknown variances. The resulting data and parameter augmentation method enable us to estimate more easily the parameters of the initial model with the help of an iterative state estimation technique for the extended model. The term Gaussian randomization is introduced for such type of modifications. The proposed method should not be confused with the Bayesian approach. The variance parameters involved in the Gaussian randomization are assumed to be unknown in contrast to the Bayesian approach in which normal priors are selected for the parameters of the model.

The estimation is performed in the classical framework of incomplete data models via an appropriate stochastic variant of an EM-type algorithm [11]. In contrast to classical stochastic variants ([6], [27], [10]) the proposed regularized particle-filter EM algorithm is particularly adapted to the extended model resulting from the Gaussian randomization, since the last iterations of the algorithm are performed under very low variance scenarios. Some recent efforts to obtain appropriate stochastic EM-type algorithms for complex models arising in plant growth model applications can be found in [25] and [26]. These algorithms are designed for a non-explicit M-step and a numerical maximization algorithm is involved as well in one of the conditional maximization steps. In this paper we present a simple way of turning a non-explicit M-step into an explicit one and opt for the post-regularized particle filter (post-RPF, e.g., [21], [23] and [16]) to perform the E-step. We mention that the use of this type of particle filter was restricted until now to Bayesian type estimation.

The method is proved to be well-adapted for mathematical models of plant growth and more generally in life sciences, for which model parameterization is generally a difficult process. Such type of models are characterized by a large number of interacting processes with a large number of model parameters, nonlinear dynamics and scarce data for parameter estimation, resulting from costly experimental data acquisition. Bayesian inference methods offer interesting perspectives to cope with such characteristics, mostly due to their robustness when confronted to scarce data, and have been used for this pur-

pose in plant growth modeling ([28], [13]). However, in such situations the influence of a priori distributions is prevalent ([7]) while it is generally not easy to assess precisely. Our approach can be seen as an effort to combine the robustness of Bayesian methods while preserving the prevalence of observation data and the interest of point estimation for mechanistic plant models, in which the biophysical meaning and value of parameters are crucial.

The paper is organized as follows. In Section 2, we introduce the notion of Gaussian randomization and present a theoretical framework for transforming the parameter estimation problem in the initial model into a problem that can be solved with the help of an appropriate EM-type algorithm in the resulting incomplete data model. In Section 3, we illustrate this idea in a toy Gaussian model, where analytic solutions are available. In Section 4, we first describe a parametric model which arises in the description of the sugar beet growth, then give a state space model representation and finally apply the idea of Gaussian randomization for this model. In Section 5, we introduce a regularized particle filter approximation of the E-step of the EM algorithm in order to make feasible parameter estimation and give a synthetic example. In order to illustrate the performance of the algorithm with real data, we present in Section 6 a case study with irregular observations (scarce data) from the sugar beet plant.

2 Gaussian randomization

Let us consider an arbitrary parametric statistical model $m \triangleq \{(\Omega, \mathcal{A}, \mathbb{P}_\mu); \mu \in M\}$, where M is a euclidean subset, and a random vector Y (representing the data vector) defined on this probability space and taking values in a measured space $\{(\mathcal{Y}, \mathcal{B}, \nu)\}$, where ν is a reference measure. In this paper the vector Y admits a density $p(y; \mu)$ w.r.t. ν for each $\mu \in M$ and for a given observation $Y = y$, the likelihood function is denoted by $L(\mu)$. The following assumption will be used in the sequel:

Assumption 1. *i) The likelihood function $L(\mu)$ is continuous on M,*
ii) The model has a unique MLE μ^*, i.e., for all $\mu \in M$,

$$0 \leq L(\mu) \leq L(\mu^*),$$

and the second inequality is strict for $\mu \neq \mu^*$.

Let us suppose that maximization of $L(\mu)$ or $\log L(\mu)$ w.r.t. some of the parameters of μ is very difficult to perform. In particular, let us decompose the unknown parameter $\mu = (\mu_1, \mu_2)$, where μ_1 could represent a part of the parameter vector that could not be updated explicitly in an iterative conditional maximization procedure by fixing μ_2. The general idea consists in considering an enlarged model, by adding parameters and hidden variables to the existing model, in such a way that maximization w.r.t. to the augmented parameter vector in the enlarged model is equivalent to the initial maximization problem. In this direction, we give a first definition. Let d_x denote the dimension of a vector x. We treat the case where $\mu_1 \in \mathbb{R}^{d_{\mu_1}}$.

Definition 1. Let m be a statistical model which satisfies Assumption 1 and $\mu = (\mu_1, \mu_2)$, where $\mu_1 \in \mathbb{R}^{d_{\mu_1}}$. The statistical model $\widetilde{m}(\mu_1)$ will be called Gaussian randomization of m w.r.t. μ_1 if $\widetilde{m}(\mu_1)$ is an incomplete data model, which consists in:

i) a Gaussian hidden vector Θ, where

$$\Theta \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma),$$

and $\Sigma = \mathrm{diag}\{\sigma_i^2\}_{1 \leq i \leq d_{\mu_1}}$, where $\sigma_i^2 > 0$,

ii) an observed vector Y, where the conditional distribution of Y given $\Theta = \theta$ depends only on the parameter μ_2 and satisfies:

$$p(y|\theta; \mu_2) = L(\theta, \mu_2).$$

Let us take $\sigma^2 = (\sigma_i^2)_{1 \leq i \leq d_{\mu_1}} \in (\mathbb{R}_+^*)^{d_{\mu_1}}$ to be a minimal vector representation of the corresponding covariance matrix Σ. The parameterization corresponding to $\widetilde{m}(\mu_1)$ is given by:

$$\phi = (\mu, \sigma^2) = (\mu_1, \mu_2, \sigma^2) \in \Phi = M \times (\mathbb{R}_+^*)^{d_{\mu_1}} \subset \mathbb{R}^{d_\phi}. \tag{1}$$

Let us now assume that we allow some variance parameters of Θ to be null. This is equivalent to say that the corresponding parameters are not randomized. Denote by μ_{11} the subvector of μ_1 with associated strictly positive variances. According to Definition 1-i) this model is no longer a Gaussian randomization of m w.r.t. μ_1 but rather a Gaussian randomization of m w.r.t. μ_{11}. In order to treat in a common framework the class of all Gaussian randomizations of m w.r.t. to subvectors of μ_1 we allow null variances in

the parameterization. Most importantly, if σ^2 is assigned to the null vector $0_{d_{\mu_1}}$, then the parameter $\phi = (\mu_1, \mu_2, 0_{d_{\mu_1}})$ could be identified with the parameter $\mu = (\mu_1, \mu_2)$ of the initial model m. In this way m could be understood as a submodel of $\widetilde{m}(\mu_1)$. In the sequel we justify theoretically this intuition. Let us first state the following lemma.

Lemma 1. *The likelihood function $\tilde{L}(\phi)$ of the extended model $\widetilde{m}(\mu_1)$ as given in Definition 1 is upper bounded by $L(\mu^*)$.*

Proof: First, we compute the likelihood function $\tilde{L}(\phi)$ of the extended model $\widetilde{m}(\mu_1)$. We have:

$$\begin{aligned} \tilde{L}(\phi) &= p(y;\phi) = \int_{\theta \in M_1} p(\theta, y; \phi) d\theta \\ &= \int_{\theta \in M_1} p(\theta; \mu_1, \sigma^2)\, p(y|\theta; \mu_2) d\theta \\ &= \mathbb{E}_{\mu_1, \sigma^2}\big(p(y|\Theta; \mu_2)\big), \end{aligned} \tag{2}$$

where $\Theta \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma)$ according to Definition 1-i). By using Definition 1-ii) and Assumption 1-ii) we get successively that for any $\phi \in \Phi$,

$$\tilde{L}(\phi) = \mathbb{E}_{\mu_1, \sigma^2}\big(L(\Theta, \mu_2)\big) \leq L(\mu^*), \tag{3}$$

and this completes the proof. □

In order to justify the existence of the MLE in this extended model, we need to extend the likelihood function of $\widetilde{m}(\mu_1)$ to parameter values which include null variances as explained at the beginning of this section and give the correct interpretation in this type of boundary values. We will need the following lemma.

Lemma 2. *Let $x \in \mathbb{R}^{d_x}$, $I \subset \{1, \ldots, d_x\}$ and $J = \{1, \ldots, d_x\} \setminus I$. Let also σ_I^2 and σ_J^2 be the subvectors of $\sigma^2 \in \mathbb{R}^{d_x}$ with components indexed by I and J respectively. If we consider the family of non-singular random vectors $\{X_{\sigma^2}; X_{\sigma^2} \sim \mathcal{N}_{d_x}(x, \Sigma)\}$, where $\Sigma = diag\{\sigma_i^2\}_{1 \leq i \leq d_x}$, then we have*

$$X_{\sigma^2} \xrightarrow[||\sigma_J^2|| \to 0]{weakly} X_{\sigma_I^2} \sim \mathcal{N}_{d_x}(x, \Sigma_I),$$

where Σ_I is a singular covariance matrix which results from Σ by putting $\sigma_J^2 = 0_J$. In particular, if $I = \emptyset$, then

$$X_{\sigma^2} \xrightarrow[||\sigma^2|| \to 0]{weakly} x.$$

Proof: We denote by ϕ_{σ^2}, $\phi_{\sigma_I^2}$ and ϕ_0 the characteristic functions of X_{σ^2}, $X_{\sigma_I^2}$ and the constant vector x respectively. It suffices to show that as $\sigma_J^2 \to 0_J$, ϕ_{σ^2} converges pointwise to $\phi_{\sigma_I^2}$. Indeed, let $t = (t_1, \ldots, t_{d_x})^\top \in \mathbb{R}^{d_x}$. We have

$$\phi_{\sigma^2}(t) = e^{it^\top x - \frac{1}{2}t^\top \Sigma t} \xrightarrow[||\sigma_J^2||\to 0]{} e^{it^\top x - \frac{1}{2}t^\top \Sigma_I t} = \phi_{\sigma_I^2}(t). \tag{4}$$

In particular, if $I = \emptyset$, then Σ_I is the null matrix and by (4), ϕ_{σ^2} converges pointwise to ϕ_0. $\square$

Proposition 1. *Let $I \subset \{1, \ldots, d_{\mu_1}\}$ and $J = \{1, \ldots, d_{\mu_1}\} \backslash I$. Let also μ_I be the subvector of μ_1 with components indexed by I, and σ_I^2, σ_J^2 and Σ_I as given in Lemma 2. We have*

$$\tilde{L}(\phi) = \tilde{L}(\mu, \sigma_I^2, \sigma_J^2) \xrightarrow[||\sigma_J^2||\to 0]{} \tilde{L}(\mu, \sigma_I^2),$$

where $\tilde{L}(\mu, \sigma_I^2)$ is the likelihood function associated with the $\widetilde{m}(\mu_I)$ Gaussian randomization of m. In particular, if $I = \emptyset$, then

$$\tilde{L}(\phi) = \tilde{L}(\mu, \sigma^2) \xrightarrow[||\sigma^2||\to 0]{} L(\mu),$$

where $L(\mu)$ is the likelihood function associated with the initial model m.

Proof: By (3) the likelihood $\tilde{L}(\phi)$ is expressed as the expectation of the random variable $L(\Theta, \mu_2)$, where $\Theta \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma)$. For clarity, let us denote $\{\Theta_{\sigma^2}\}$ the family of random vectors Θ indexed by σ^2. Then, we can rewrite (3) as

$$\tilde{L}(\phi) = \mathbb{E}_{\mu_1, \sigma^2}\big(L(\Theta_{\sigma^2}, \mu_2)\big). \tag{5}$$

If we apply Lemma 2 for the family $\{\Theta_{\sigma^2}\}$, we have that for any (μ_1, σ_I^2),

$$\Theta_{\sigma^2} \xrightarrow[||\sigma_J^2||\to 0]{\text{weakly}} \Theta_{\sigma_I^2} \sim \mathcal{N}_{d_{\mu_1}}(\mu_1, \Sigma_I),$$

or equivalently, for any continuous and bounded function $f : \mathbb{R}^{d_{\mu_1}} \to \mathbb{R}$,

$$\mathbb{E}_{\mu_1, \sigma^2}\big(f(\Theta_{\sigma^2})\big) \xrightarrow[||\sigma_J^2||\to 0]{} \mathbb{E}_{\mu_1, \sigma_I^2, 0_J}\big(f(\Theta_{\sigma_I^2})\big), \tag{6}$$

where we have used the characterization of weak convergence of a family of random vectors (indexed by a continuous parameter) to a random vector in a specific limit point.

By Assumption 1, the likelihood function $L(\mu)$ is continuous on M and bounded by $L(\mu^*)$. Consequently, for any μ_2, the function $g : \mathbb{R}^{d_{\mu_1}} \to \mathbb{R}$, where $\mu_1 \mapsto L(\mu_1, \mu_2)$ is continuous on M_1 and bounded. Using this and (6), we have

$$\mathbb{E}_{\mu_1,\sigma^2}\big(L(\Theta_{\sigma^2}, \mu_2)\big) = \mathbb{E}_{\mu_1,\sigma^2}\big(g(\Theta_{\sigma^2})\big) \xrightarrow[||\sigma_J^2||\to 0]{} \mathbb{E}_{\mu_1,\sigma_I^2,0_J}\big(g(\Theta_{\sigma_I^2})\big) = \mathbb{E}_{\mu_I,\sigma_I^2}\big(L(\Theta_{I,\sigma_I^2}, \mu_J, \mu_2)\big), \tag{7}$$

where Θ_{I,σ_I^2} is the subvector of Θ_{σ^2} indexed by I. In the last expectation $\sigma_i^2 > 0$ only for $i \in I$ and consequently the last term corresponds to the analogue of (5) for an equivalent representation of the likelihood function associated to the $\widetilde{m}(\mu_I)$ Gaussian randomization of m. If $I = \emptyset$, then obviously the last term coincides with $L(\mu)$ and the proof is complete. □

By Proposition 1 we can consider in a unified framework the class $\{\widetilde{m}(\mu_I)\}_I$ of $2^{d_{\mu_1}} - 1$ Gaussian randomizations of m w.r.t. to subvectors of μ_1 together with the initial model m, which corresponds to the choice $I = \emptyset$. For this reason we extend the domain of $\tilde{L}(\phi)$, from Φ given by (1) to $\tilde{\Phi} = M \times (\mathbb{R}_+)^{d_{\mu_1}}$ by allowing null variances as follows:

$$\tilde{L}(\mu, \sigma_I^2, 0_J) = \tilde{L}(\mu, \sigma_I^2), \tag{8}$$

$$\tilde{L}(\mu, 0_{d_{\mu_1}}) = L(\mu). \tag{9}$$

With similar arguments as in the proof of Proposition 1 we can deduce from the continuity of $L(\mu)$ on M that $\tilde{L}(\phi)$ is continuous on Φ (as $(\mu, \sigma^2) \to (\mu_0, \sigma_0^2) \in \Phi$ show that $(\Theta_{\mu_1,\sigma^2}, \mu_2)$ converges weakly to $(\Theta_{\mu_{1,0},\sigma_0^2}, \mu_{2,0})$). Consequently, both equations (8) and (9) determine the continuous extension of $\tilde{L}$ from Φ to $\tilde{\Phi}$ by using Proposition 1.

Theorem 1. *Let us consider the class $\widetilde{m}_{\mu_1}$ of all Gaussian randomizations $\{\widetilde{m}(\mu_I)\}_I$ of the initial model m, where $I \subset \{1, \ldots, d_{\mu_1}\}$ (for $I = \emptyset$, $\widetilde{m}(\mu_\emptyset) = m$) and μ_I is defined in Proposition 1. If Assumption 1 holds, then the unique MLE μ^* associated with the model m determines a unique MLE ϕ^* associated with the model $\widetilde{m}_{\mu_1}$. In particular, we have*

$$\phi^* = (\mu^*, 0_{d_{\mu_1}}), \tag{10}$$

$$\tilde{L}(\phi^*) = L(\mu^*). \tag{11}$$

Conversely, if Assumption 1-i) holds and the MLE ϕ^* associated with the model $\widetilde{m}_{\mu_1}$ exists and is unique, then it determines a unique MLE μ^* associated with the model m and satisfies (10) and (11).

Proof: $\Rightarrow$) The restriction of $\tilde{L}(\phi)$ on $\Phi = M \times (\mathbb{R}_+^*)^{d_{\mu_1}}$ is upper bounded by $L(\mu^*)$ by Lemma 1. Now, if $\phi \in \tilde{\Phi} \setminus \Phi$, take the nonempty set $J = \{j \in \{1, \ldots, d_{\mu_1}\} : \sigma_j^2 = 0\}$ and I the complement of J. Then, $\tilde{L}(\phi)$ can be expressed as the left-hand member of equations (8) or (9). By the same equations we obtain that $\tilde{L}(\phi)$ equals the likelihood function of the extended model $\tilde{m}(\mu_I)$ or of the initial model m respectively. Therefore, by applying Lemma 1 for μ_I or by Assumption 1 respectively, we infer that $\tilde{L}(\phi)$ is upper bounded by $L(\mu^*)$ also on $\tilde{\Phi} \setminus \Phi$, and consequently on $\tilde{\Phi}$. By (9) we have that $\tilde{L}(\mu^*, 0_{d_{\mu_1}}) = L(\mu^*)$ and consequently the MLE exists and if it is unique it satisfies (10) and (11). We have that $(\mu^*, 0_{d_{\mu_1}}) \in M \times \{0\}^{d_{\mu_1}}$ and by Assumption 1 and (9), it is unique on this set. Let us now assume that there exists $\phi = (\mu, \sigma^2) \in M \times \left(\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}}\right)$ such that $\tilde{L}(\mu, \sigma^2) = L(\mu^*)$. Take the nonempty set $I = \{i \in \{1, \ldots, d_{\mu_1}\} : \sigma_i^2 > 0\}$ and J the complement of I. By (8), (3), the remark following (7) and our assumption, we have

$$\tilde{L}(\mu, \sigma^2) = \mathbb{E}_{\mu_I, \sigma_I^2}\big(L(\Theta_I, \mu_2^{'})\big) = L(\mu^*), \tag{12}$$

where $\mu_2^{'}$ concatenates the components of μ_1 indexed by J and μ_2. Since by Assumption 1, the function $L(\mu)$ is upper bounded by $L(\mu^*)$, we obtain that the random variable $L(\Theta_I, \mu_2^{'})$ has the same upper bound, that is, $L(\mu^*) - L(\Theta_I, \mu_2^{'})$ is a positive random variable. But, by (12)

$$\mathbb{E}_{\mu_I, \sigma_I^2}\big(L(\mu^*) - L(\Theta_I, \mu_2^{'})\big) = 0,$$

so we get $L(\Theta_I, \mu_2^{'}) = L(\mu^*)$, $\mathbb{P}_{\mu_I, \sigma_I^2}-$a.s. , or

$$1 = \mathbb{P}_{\mu_I, \sigma_I^2}\big(L(\Theta_I, \mu_2^{'}) = L(\mu^*)\big) = \mathbb{1}_{\{\mu_2^{'} = \mu_2^{'*}\}} \mathbb{P}_{\mu_I, \sigma_I^2}(\Theta_I = \mu_I^*) = 0,$$

where the second equality follows by Assumption 1 and the last equality follows from the fact that Θ_I follows a non-singular multivariate normal distribution, so it induces on $(\mathbb{R}^{d_{\mu_I}}, \mathcal{B}(\mathbb{R}^{d_{\mu_I}}))$ a measure which is absolutely continuous w.r.t. the Lebesgue measure which assigns measure 0 to the singleton $\{\mu_I^*\}$. We conclude that for any $\phi \in M \times \left(\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}}\right)$, we have $\tilde{L}(\mu, \sigma^2) < L(\mu^*)$ and consequently the MLE ϕ^* is unique.

$\Leftarrow$) If an MLE exists on $\tilde{\Phi}$, then $\tilde{L}$ is upper bounded on $\tilde{\Phi}$ by $\tilde{L}(\phi^*)$, and consequently is upper bounded on $M \times \{0\}^{d_{\mu_1}}$ by the same value. By (9) this implies that L is upper bounded on M by $\tilde{L}(\phi^*)$. If $\phi^* \in M \times \{0\}^{d_{\mu_1}}$, then (10) and (11) are satisfied for some $\mu^* \in M$. Since the upper bound of L is attained on μ^* the MLE exists. It is also unique, since if μ_1^* and μ_2^* are two distinct MLE, then $(\mu_1^*, 0_{d_{\mu_1}})$ and $(\mu_2^*, 0_{d_{\mu_1}})$ are two distinct MLE

associated with $\widetilde{m}_{\mu_1}$, and this contradicts the initial assumption. Now we will show that $\phi^* \in M \times \{0\}^{d_{\mu_1}}$ is a necessary condition. Let us assume that $\phi^* \in M \times \Big(\mathbb{R}_+^{d_{\mu_1}} \setminus \{0\}^{d_{\mu_1}}\Big)$. Take the nonempty set $I = \{i \in \{1, \ldots, d_{\mu_1}\} : (\sigma_i^2)^* > 0\}$ and J the complement of I. By applying (12) for ϕ^* we have

$$\tilde{L}(\phi^*) = \mathbb{E}_{\mu_I^*, (\sigma_I^2)^*}\big(L(\Theta_I, (\mu_2^{'})^*)\big). \tag{13}$$

If $L(\mu_I, (\mu_2^{'})^*) < \tilde{L}(\phi^*)$ for all $\mu_I \in \mathbb{R}^{d_{\mu_I}}$ then this would contradict (13). So, we conclude that there exists μ_I^{**} such that $L(\mu_I^{**}, (\mu_2^{'})^*) = \tilde{L}(\phi^*)$, or $\phi^{**} = (\mu_I^{**}, (\mu_2^{'})^*, 0_{d_{\mu_1}}) \neq \phi^*$ is an MLE, which contradicts the assumption of uniqueness. Therefore, $\phi^* \in M \times \{0\}^{d_{\mu_1}}$ and the proof is complete. □

A consequence of Theorem 1 is that we can transfer the maximization problem corresponding to the initial model m to a maximization problem corresponding to the model $\widetilde{m}_{\mu_1}$. But since the latter is formulated as an incomplete data problem, we could design an appropriate version of the EM-algorithm to solve the maximization problem or more often a stochastic variant of the EM-algorithm (Expectation-Maximization, see [11]) to handle the usually non-explicit state estimation problem. It is noteworthy that if all parameters belonging to μ are randomized then the M-step becomes explicit. Otherwise, a GEM (G:Generalized) or stochastic GEM could be used to find the MLE. An example of this kind could be an ECM (C:Conditional), see [20], where parameters are updated in a cyclic fashion. In the following Proposition we give the form of the Q-function, and the update formulas for the randomized parameter μ_1.

Proposition 2. *Let $\phi' = (\mu_1, \mu_2, \sigma^2)' \in M \times (\mathbb{R}_*^+)^{d_{\mu_1}}$ be the current parameter update associated with the model $\widetilde{m}_{\mu_1}$. The update equations for the parameter (μ_1, σ^2) are given by:*

$$\widehat{\mu_1} = \mathbb{E}_{\phi'}\big(\Theta \mid Y = y\big) \tag{14}$$

$$\widehat{\sigma^2} = (\widehat{\sigma_i^2})_{1 \le i \le d_{\mu_1}} = \big(Var_{\phi'}(\Theta_i \mid Y = y)\big)_{1 \le i \le d_{\mu_1}} \tag{15}$$

and μ_2 can be updated independently by maximizing

$$Q_2(\mu_2; \phi^{'}) = \mathbb{E}_{\phi'}\big\{\log p(y|\Theta; \mu_2) \mid Y = y\big\}. \tag{16}$$

Proof: Since $\phi' \in M \times (\mathbb{R}_+^*)^{d_{\mu_1}}$, this parameter corresponds to the Gaussian randomization $\widetilde{m}(\mu_1)$ and by the definition of the Q-function and Definition 1 we have

$$\begin{aligned} Q(\phi;\phi') &= \mathbb{E}_{\phi'}\big\{\log p(\Theta, y; \mu_1, \mu_2, \sigma^2) \mid Y = y\big\} && (17)\\ &= \mathbb{E}_{\phi'}\big\{\log p(\Theta; \mu_1, \sigma^2) \mid Y = y\big\} + \mathbb{E}_{\phi'}\big\{\log p(y|\Theta; \mu_2) \mid Y = y\big\}. && (18) \end{aligned}$$

The last term in the right-hand member of equation (18) depends only on μ_2 and coincides with (16), while the first term denoted by $Q_1(\mu_1, \sigma^2; \phi')$ does not depend on μ_2. Consequently, the last statement of this theorem is true. By using the independence of the components of Θ and the density of the normal distribution we have that Q_1 can be maximized equivalently as follows:

$$Q_1(\mu_1, \sigma^2; \phi') \overset{\max}{\sim} -\frac{1}{2}\sum_{i=1}^{d_{\mu_1}} \log \sigma_i^2 - \frac{1}{2}\sum_{i=1}^{d_{\mu_1}} \frac{\mathbb{E}_{\phi'}\{(\Theta_i - \mu_{1,i})^2 \mid Y = y\}}{\sigma_i^2}. \qquad (19)$$

Consequently, it is elementary to show by relation (19) that (14) and (15) hold. □

Remark 1. A question which arises naturally concerns the adaptation of this method in the case that some of the components of the parameter μ have range different than $\mathbb{R}$. Then a Gaussian randomization as explained in Definition 1 cannot be applied directly and a modification is needed by suitable reparameterizations whenever possible. Examples of this kind are given in subsection 4.3.

3 A toy example of Gaussian randomization

In order to illustrate the proposed method, we start this section by a simple example where the MLE is known. In the next section we consider an application to a model which arises from plant growth modeling and the MLE has no explicit form.

Let us consider a simple Gaussian model. In particular, take a vector $Y = y$ of n i.i.d. observations $\{Y_i = y_i\}_{i=1}^n$ from a univariate normal distribution $\mathcal{N}(\mu_y, \sigma_y^2)$, where $\mu_y \in \mathbb{R}$ is an unknown parameter, $\sigma_y^2 \in \mathbb{R}_+^*$. We study the effect of Gaussian randomization when i) σ_y^2 is known and ii) σ_y^2 should be estimated.

Let us first treat case i). Since the initial model m is Gaussian it is well-known that the MLE exists, is unique and is given by the sample mean, that is, $\mu_y^* = \bar{y}$.

Now assume that a Gaussian randomization is introduced for the parameter μ_y. In particular, set $\mu = \mu_y$ and consider the Gaussian randomization $\widetilde{m}(\mu)$. In this case,

$$Y \sim \mathcal{N}_n\big(\Theta 1_n, \sigma_y^2 I_n\big), \quad \Theta \sim \mathcal{N}(\mu, \sigma^2), \tag{20}$$

where 1_n is the column-vector of ones and I_n is the identity matrix of rank n. By writing,

$$Y_i = \mu + \sigma\xi + \sigma_y \epsilon_i, \quad i = 1, \ldots, n,$$

where $\{\xi, \epsilon_i \,; i = 1, \ldots, n\}$ are mutually independent standard Gaussian random variables, we get easily that

$$Y \sim \mathcal{N}_n\big(\mu 1_n, \Sigma_{\sigma^2}\big), \tag{21}$$

where the elements of the covariance matrix Σ_{σ^2} are given by

$$\Sigma_{\sigma^2}(i,j) = \begin{cases} \sigma^2 + \sigma_y^2 & \text{if } i = j, \\ \sigma^2 & \text{if } i \neq j. \end{cases} \tag{22}$$

Notice that in the extended model $\widetilde{m}(\mu)$ the r.v. $\{Y_i\}_{i=1}^n$ are no longer independent. Now, we take $\phi = (\mu, \sigma^2)$, $\Phi = \mathbb{R} \times \mathbb{R}_+^*$ and we write the equivalent log-likelihood in the sense of maximization

$$\tilde{L}(\phi) \overset{\max}{\sim} -\frac{1}{2}\log\det\Sigma_{\sigma^2} - \frac{1}{2}\,(y - \mu 1_n)^\top\,\Sigma_{\sigma^2}^{-1}\,(y - \mu 1_n), \quad \phi \in \Phi. \tag{23}$$

By Theorem 1, we should have that a continuous extension of $\tilde{L}(\phi)$ from Φ to $\tilde{\Phi} = \mathbb{R} \times \mathbb{R}_+$ is feasible, the MLE exists in the extended model, is unique and is given by $\phi^* = (\mu^*, 0)$. As far as the extension is concerned, we verify by (23) that $\lim_{\sigma^2 \to 0} \tilde{L}(\phi) = L(\mu)$, where $L(\mu)$ is the likelihood function of the initial model m, and consequently a continuous extension is possible. In order to verify the aforementioned properties of the MLE, it is advantageous to use the simple form of the covariance matrix Σ_{σ^2} given by (22). We can easily get that

$$\det\Sigma_{\sigma^2} = (\sigma_y^2)^{n-1}(n\sigma^2 + \sigma_y^2), \qquad n \geq 1, \tag{24}$$

and also that

$$\Sigma_{\sigma^2}^{-1}(i,j) = \begin{cases} \dfrac{1}{\sigma_y^2}\left(1 - \dfrac{\sigma^2}{n\sigma^2 + \sigma_y^2}\right) & \text{if } i = j, \quad n \geq 1, \\ -\dfrac{\sigma^2}{\sigma_y^2(n\sigma^2 + \sigma_y^2)} & \text{if } i \neq j, \quad n \geq 2. \end{cases} \tag{25}$$

Now, by inserting (24) and (25) in (23), we get

$$\tilde{L}(\phi) \overset{\max}{\sim} -\frac{1}{2}\log(n\sigma^2+\sigma_y^2) - \frac{1}{2\sigma_y^2}\sum_{i=1}^{n}(y_i-\mu)^2 + \frac{\sigma^2}{2\sigma_y^2(n\sigma^2+\sigma_y^2)}\left\{\sum_{i=1}^{n}(y_i-\mu)\right\}^2. \quad (26)$$

By taking the partial derivative of (26) w.r.t. μ we get

$$\partial\tilde{L}(\phi)/\partial\mu = \frac{1}{n\sigma^2+\sigma_y^2}\sum_{i=1}^{n}(y_i-\mu). \quad (27)$$

By setting (27) equal to zero, we get indeed as unique solution μ^* the sample mean $\bar{y}$. By inserting this solution into (26) the third term of the right-hand member of (26) gets null. So, the maximization w.r.t. σ^2 is equivalent to the maximization of the first term, which is a strictly decreasing function of σ^2 and consequently attains its maximum at $\sigma^2 = 0$. This shows that $\phi^* = (\mu^*, 0)$ as expected. Conversely, and this is the most interesting direction in applications, since the MLE has the form $(\mu^*, 0)$ in the extended model, we conclude by Theorem 1 that μ^* is the MLE in the initial model.

It is noteworthy that in interesting problems that this idea can be applied, the direct maximization problem in the extended model will in general be no easier than the maximization in the initial model. But, its formulation as a hidden variable model gives directly a recursive procedure via an appropriate version of the EM algorithm to estimate the parameters of the model. In this example the M-step is explicit since the conditional distribution of Θ given Y is normal. Let ϕ' denote the current update of ϕ. We obtain easily that

$$[\Theta|Y=y] \sim \mathcal{N}\left(\frac{n(\sigma^2)'\,\bar{y}+\sigma_y^2\,\mu'}{n(\sigma^2)'+\sigma_y^2}, \frac{(\sigma^2)'\sigma_y^2}{n(\sigma^2)'+\sigma_y^2}\right). \quad (28)$$

Consequently, by Proposition 2, the update equations for μ and σ^2 are given by the mean and the variance of this normal distribution respectively. Note also that the likelihood of the extended model given by (23) has a unique stationary point which coincides with the MLE, so the sequence of EM iterates given by (28) is guaranteed to converge to the MLE.

Let us now treat case ii). The previous analysis shows that the MLE of (μ_y, σ^2) does not involve the parameter σ_y^2 and for this reason the MLE of μ_y and σ^2 are identical with those of case i). After maximizing $\tilde{L}(\phi)$ given by (23) with respect to (μ_y, σ^2), the MLE

of σ_y^2 can be obtained by maximizing the function

$$f(\sigma_y^2) = -\frac{n}{2}\log(\sigma_y^2) - \frac{1}{2\sigma_y^2}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

with respect to σ_y^2. This gives that $(\sigma_y^2)^* = n^{-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, that is, the sample variance, which obviously coincides with the MLE of the initial model of i.i.d. normal observations. The point here is that these parameters that conditional on the others can be solved explicitly (as function of the others) in the initial likelihood formulation, they still keep this advantage when estimated with the EM algorithm. Indeed, in this example the second term of the Q-function given by (16) can be written as:

$$Q_2(\sigma_y^2; \phi') \overset{\max}{\sim} -\frac{n}{2}\log(\sigma_y^2) - \frac{1}{2\sigma_y^2}\sum_{i=1}^{n}\mathbb{E}_{\phi'}\left\{(y_i - \Theta)^2 \mid y\right\}. \tag{29}$$

By maximizing the above quantity we get:

$$\widehat{\sigma}_y^2 = g(\widehat{\mu}_y, \widehat{\sigma}^2) = n^{-1}\sum_{i=1}^{n} y_i^2 - \widehat{\mu}_y\,(2\bar{y} - \widehat{\mu}_y\,) + \widehat{\sigma}^2, \tag{30}$$

where $\widehat{\mu}_y$ and $\widehat{\sigma}^2$ are the current updates of μ_y and σ^2 given by (28), and since $\widehat{\mu}_y \to \mu^*$ and $\widehat{\sigma}^2 \to 0$, we conclude by (30) that $\widehat{\sigma}_y^2$ converges to $(\sigma_y^2)^*$.

Remark 2. It is noteworthy that the addition of these artificial parameters does not affect the MLE of the initial model, even if one could think at first sight that the extended model is not identifiable. For example, consider the extreme and trivial case where $n = 1$ and σ_y^2 is known. We have one available observation and two parameters μ_y and σ^2 to estimate. But in fact here, the subsequent degeneration of $\widehat{\sigma}^2$ to zero, which we know that it is the true value, causes no problem. On the contrary, notice the difference if one wished to estimate in the initial model μ_y and σ_y^2 jointly (a truly non identifiable model). We have

$$L(\mu_y, \sigma_y^2) \overset{\max}{\sim} (\sigma_y^2)^{-1/2}\exp\left\{\frac{-(y-\mu_y)^2}{2\sigma_y^2}\right\}, \quad \mu_y \in \mathbb{R},\ \sigma_y^2 > 0, \tag{31}$$

$$\lim_{\sigma_y^2 \to 0} L(\mu_y, \sigma_y^2) = \begin{cases} +\infty & \text{if } \mu_y = y, \\ 0 & \text{if } \mu_y \neq y. \end{cases} \tag{32}$$

We deduce that the likelihood function is unbounded and the MLE does not exist. Additionally, it does not admit any continuous extension to be able to approach from the larger model the obvious solution $\mu_y = y$ when μ_y is the only parameter to be estimated.

4 An application to a plant growth model

Let us consider a nonlinear dynamic stochastic model of individual plant growth, the LNAS (Log-Normal Allocation and Senescence) model [8], whose aim is to describe the growth of the sugar beet plant in terms of the evolution of its biomass (mass for living organisms) with a daily time step. This model can therefore be regarded as a simplified version of the GreenLab model [18] which is more adapted to experimental trials when only compartment data are available.

4.1 Description of the model

In this example, we present the case where the structure of the sugar beet can be simplified as the set of two types of organ compartments, the leaves and the root, from now on labeled as compartment 1 and 2 respectively. The development of sugar beet can be described via three basic mechanisms, biomass production, biomass allocation to these compartments and leaf senescence.

First, we describe the procedure of biomass production. The initial mass of the plant, corresponds to the mass of the seed and is denoted by q_0. The subsequent biomasses $q_n,\ n \geq 1$, result from photosynthesis and under no modeling errors are assumed to be given by an adaptation of the empirical Beer-Lambert law [14],

$$q_n = \mu_a \cdot \mathrm{PAR}_n \cdot \big(1 - \exp\{-\lambda\, q_n^g\}\big), \tag{33}$$

where μ_a is a strictly positive biological parameter related to the radiation use efficiency, q_n^g is the total mass of green leaves at day n (leaves that are still (photosynthetically active), PAR_n is the photosynthetically active radiation at the n-th day and λ an empirical parameter. In [18], λ is decomposed in the product of the Beer-Lambert extinction coefficient k by the specific leaf area divided by S_p, a characteristic surface related to inter-individual competition.

In the sequel, for notational simplicity, we denote by F_n the function given by (33) which relates q_n with q_n^g by omitting the dependence on the other variables. In order to give some flexibility to this model and to explain some observed variability around the theoretical values given by (33), we consider a modified model of the form:

$$Q_n = F_n(Q_n^g; \mu_a, \lambda)\, e^{\xi_n^q}, \qquad n \geq 1, \tag{34}$$

where Q_n and Q_n^g are the stochastic counterparts of q_n and q_n^g respectively, related via the function F_n given by (33), and $\{\xi_n^q\}_{n\geq 1}$ is an independent sequence with $\xi_n^q \sim \mathcal{N}(0, \sigma_q^2)$.

Now, we describe the procedure of biomass allocation. Since we assume two organ compartments, the allocation of each produced biomass Q_n is divided into two parts, one for the leaves compartment and one for the root. Let $G(x; \mu, \sigma^2)$ denote the distribution function of the log-normal distribution with location and scale parameters given by μ and σ respectively, and $G^c(x; \mu, \sigma^2)$ its corresponding survival function, that is, $G^c(x) = 1 - G(x)$. Let also τ be the cumulative daily temperature (thermal time in ℃.*days*) and define the function

$$\gamma(\tau; \gamma_0, \gamma_f, \mu_\gamma, \sigma_\gamma^2) = \gamma_0 + (\gamma_f - \gamma_0)\, G(\tau; \mu_\gamma, \sigma_\gamma^2), \qquad \tau > 0, \tag{35}$$

where γ_0, γ_f are parameters such that $0 < \gamma_0 < \gamma_f < 1$. The allocated percentage to compartment 1 (leaves) on day n associated with thermal time τ_n, denoted by Γ_n, is assumed to be a logit-normal distribution with mean the logit transform of $\gamma(\tau_n; \gamma_0, \gamma_f, \mu_\gamma, \sigma_\gamma^2))$ and variance $\sigma_{\gamma\gamma}^2$, that is,

$$\log\left(\frac{\Gamma_n}{1-\Gamma_n}\right) = \log\left(\frac{\gamma(\tau_n; \mu_\gamma, \sigma_\gamma^2)}{1-\gamma(\tau_n; \mu_\gamma, \sigma_\gamma^2)}\right) + \xi_n^\gamma, \tag{36}$$

where $\xi_n^\gamma \sim \mathcal{N}(0, \sigma_{\gamma\gamma}^2)$. In this way, the total leaf mass and the total root mass can be computed recurrently by:

$$Q_{n+1}^f = Q_n^f + \Gamma_n\, Q_n, \tag{37}$$

$$Q_{n+1}^r = Q_n^r + (1-\Gamma_n) Q_n, \tag{38}$$

where the initial conditions $Q_0^f = Q_0^r = 0$ hold, Q_n and Γ_n satisfy (34) and (36) respectively.

The last important process to consider is leaf senescence. We model with the following equation the senescent leaf mass Q_n^s until day n, which corresponds to the part of the total leaf mass at day n that is no longer photosynthetically active (it may have already fallen from the plant or is still on the plant but yellow). We assume that

$$Q_n^s = G(\tau_n; \mu_s, \sigma_s^2)\, Q_n^f. \tag{39}$$

By (39) we deduce directly that the total green leaf mass is given by:

$$Q_n^g = Q_n^f - Q_n^s = G^c(\tau_n; \mu_s, \sigma_s^2)\, Q_n^f. \tag{40}$$

The potentially available data consist in the compartment mass measurements $\{Q_n^g, Q_n^r\}_{n=1}^N$, where N is a given time horizon, observed under measurement errors. Let us denote $Z_{n,1}$ and $Z_{n,2}$ the measurements of Q_n^g and Q_n^r respectively corrupted by error. In particular, we assume that for $n = 1, \ldots, N$,

$$Y_n = (\log Z_{n,j})_j = \begin{pmatrix} \log Q_n^g + \epsilon_n^g \\ \log Q_n^r + \epsilon_n^r \end{pmatrix} = \begin{pmatrix} \log\left(G^c(\tau_n; \mu_s, \sigma_s^2) Q_n^f\right) + \epsilon_n^g \\ \log Q_n^r + \epsilon_n^r \end{pmatrix}, \tag{41}$$

where $\{\epsilon_n^g, \epsilon_n^r\}_{n\geq 1}$ are mutually independent and $\epsilon_n^g \sim \mathcal{N}(0, \sigma_g^2)$, $\epsilon_n^r \sim \mathcal{N}(0, \sigma_r^2)$.

We are now in the position to represent the model given by equations (34)-(41) as a state space model.

4.2 State space model representation

A possible choice of the state vector of the model is $X_n = (Q_n^f, Q_n^r)$. In order to have the state and the observation equation of the model we need the following representation:

$$X_{n+1} = \Phi_n(X_n, \xi_n; \mu), \tag{42}$$

$$Y_n = R_n(X_n, \epsilon_n; \mu), \tag{43}$$

where $X_n = (X_{n,1}, X_{n,2}) = (Q_n^f, Q_n^r)$ and $\{\xi_n\}_{n\geq 1}$ and $\{\epsilon_n\}_{n\geq 1}$ are appropriately chosen sequences of random vectors. The parameter vector of our model is denoted by μ and includes all structural parameters $\{\mu_a, \lambda, \mu_\gamma, \sigma_\gamma^2, \gamma_0, \gamma_f\}$ eventually together with some of error variances $\{\sigma_{\gamma\gamma}^2, \sigma_q^2, \sigma_g^2, \sigma_r^2\}$. In the following lemma we explicit the state equation and the observation equation corresponding to our model.

Lemma 3. *The model given by equations (34)-(41) can be described as a state space model. The state equation is given by the following system of equations:*

$$X_{n+1,1} = \Phi_{n,1}(X_n, \xi_n; \mu) = X_{n,1} + \frac{F_n\left(G^c(\tau_n; \mu_s, \sigma_s^2)\, X_{n,1}; \mu_a, \lambda\right) e^{\xi_{n,1}}}{1 + \left(\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1\right) e^{-\xi_{n,2}}} \tag{44}$$

$$X_{n+1,2} = \Phi_{n,2}(X_n, \xi_n; \mu) = X_{n,2} + \frac{F_n\left(G^c(\tau_n; \mu_s, \sigma_s^2)\, X_{n,1}; \mu_a, \lambda\right) e^{\xi_{n,1}}}{1 + \left(\gamma^{-1}(\tau_n; \mu_\gamma, \sigma_\gamma^2) - 1\right)^{-1} e^{\xi_{n,2}}}, \tag{45}$$

where $\{\xi_n\}_{n\geq 1}$ is an independent from X_0 sequence of mutually independent bi-dimensional centered Gaussian random vectors with diagonal covariance matrix and variances σ_1^2 and

σ_2^2 *for* $X_{n,1}$ *and* $X_{n,2}$ *respectively. The observation equation is given by*

$$Y_n = R_n(X_n, \epsilon_n) = \begin{pmatrix} \log G^c(\tau_n; \mu_s, \sigma_s^2) + \log X_{n,1} \\ \log X_{n,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{n,1} \\ \epsilon_{n,2} \end{pmatrix}, \tag{46}$$

where $\{\epsilon_n\}_{n\geq 1}$ *is an i.i.d. sequence of* 2*-dimensional centered Gaussian random vectors with diagonal covariance matrix, variances* σ_i^2 *for each* $\epsilon_{n,i}$, $i = 1, 2$, *and assumed to be independent from* $\{\xi_n\}_{n\geq 1}$.

Proof: Due to the choice of the state vector $X_n = (Q_n^f, Q_n^r)$ and equations (37) and (38), we only have to show that $\Gamma_n Q_n$ and $(1 - \Gamma_n)Q_n$ coincide with the second summand of the right-hand member of (44) and (45) respectively. In both terms the numerator is common and equals Q_n. Indeed, this can be deduced by (34) by substituting Q_n^g with its expression in (40). Let us now denote by $\phi : (0, 1) \times \mathbb{R} \to \mathbb{R}$ such that $\phi(x, y) = \log(x/1 - x) + y$. Note by (36) that the left-hand member of this equation corresponds to the logit transformation of Γ_n, which is the inverse of the standard logistic sigmoid function. We deduce that

$$\Gamma_n = \frac{1}{1 + e^{-\phi(\gamma(\tau_n), \xi_{n,2})}} = \frac{1}{1 + \left(\gamma^{-1}(\tau_n) - 1\right) e^{-\xi_{n,2}}}, \tag{47}$$

and

$$1 - \Gamma_n = \frac{1}{1 + e^{\phi(\gamma(\tau_n), \xi_{n,2})}} = \frac{1}{1 + \left(\gamma^{-1}(\tau_n) - 1\right)^{-1} e^{\xi_{n,2}}}. \tag{48}$$

By multiplying both members of (47) and (48) with Q_n given by (34) and substituting these expressions in the state equations we obtain (44) and (45). The observation equation (46) of the model can be deduced directly by (41). □

4.3 Gaussian randomization of the plant growth model

In this part we apply the ideas developed in Section 2 for the state space model described in Lemma 3. Assume for simplicity of presentation that all error variances $\{\sigma_{\gamma\gamma}^2, \sigma_q^2, \sigma_g^2, \sigma_r^2\}$ are known or a priori fixed to some values. The interest will be focused on the estimation of the structural parameters and let $\mu = (\mu_a, \lambda, \mu_\gamma, \sigma_\gamma, \gamma_0, \gamma_f)$. Note also that this state space model can equivalently be described as a non-homogeneous hidden Markov model [5], and usual estimation techniques are based on some variant of the EM algorithm, by

combining an approximation of the E-step and a usually explicit M-step. Nevertheless, we can easily see that even in the complete log-likelihood of this model there is no explicit solution for μ, therefore leading to a non-explicit M-step. Otherwise, conditional maximization approaches should be invoked and this often leads to complicated generalized EM-type algorithms (see, eg., [25]), where a numerical maximization procedure should be implemented as well. It is also well known (see, eg., [19]) that the EM-type sequences do not necessarily converge to the true MLE (if it exists). In typical plant growth models with a large number of parameters, it is very difficult to ensure the existence of the MLE and the subsequent convergence of the EM iterates to the true MLE. The best that we can hope is to have a robust algorithm and the solutions to the initial maximization problem to be biologically relevant. For this reason, we take Assumption 1 for granted.

Let us first consider a suitable reparameterization of the state space model m in such a way that a Gaussian randomization can be applied. Let $\mu = (\mu_a, \lambda, \mu_\gamma, \sigma_\gamma, \gamma_0, \gamma_f)$. By Remark 1 and since $\mu_a, \sigma_\gamma^2 \in (0, +\infty)$ we deduce that a possible reparameterization can be given by $\eta = (\log \mu_a, \lambda, \mu_\gamma, \log \sigma_\gamma^2, \gamma_0, \gamma_f)$, where the range of η is $\mathbb{R}^6$. The functional parameters μ_s and σ_s which are related to the senescent process can be estimated only based on the senescent leaf mass and thus are fixed in our study case. The results of Section 2 are now directly applicable to η. In particular, let $\widetilde{m}(\eta)$ the Gaussian randomization of m w.r.t. η. By Definition 1 we have that $\widetilde{m}(\eta)$ is an incomplete data model which consists in:

i) a Gaussian hidden vector Θ, where

$$\Theta \sim \mathcal{N}_6(\eta, \Sigma),$$

and $\Sigma = \mathrm{diag}\{\sigma_i^2\}_{1 \le i \le 6}$, where $\sigma_i^2 > 0$,

ii) an observed vector Y, where conditioned on $\Theta = \theta$ is a state space model which satisfies (42) and (43), with the parameter μ replaced by θ.

The parameter of the extended model is given by $\phi = (\eta, \sigma^2) \in \mathbb{R}^6 \times (\mathbb{R}_+^*)^6$. By Proposition 2, the problem of parameter estimation of ϕ is transformed in an iterative state estimation problem. Starting with a fixed value $\phi^{(0)}$, the EM-update equations given by (14) and (15) produce a sequence $\phi^{(n)}$ which converges to a stationary point of the initial likelihood.

Unfortunately, the conditional moments involved in these equations cannot be computed explicitly for nonlinear models of this type and for this reason we need to implement a stochastic variant of the EM-algorithm to approximate them. In order to tackle this problem, we describe in the next section an appropriate particle filtering method.

5 Particle filtering approximation of the E-step

At each iteration of the EM-algorithm, given the current parameter update ϕ', our objective is to approximate $\mathbb{E}_{\phi'}\big(\Theta \mid Y_{0:n} = y_{0:n}\big)$ and $\big(\mathrm{Var}_{\phi'}(\Theta_i \mid Y_{0:n} = y_{0:n})\big)_{1\leq i\leq d_{\mu_1}}$ where n corresponds to the observation length, since the update equations for the parameters of the model given in Proposition 2 generally lead to non-explicit solutions. This is also the case for the specific application that we present in the previous section. One could think of several alternative algorithms to make this approximation, like Markov chain Monte-Carlo or particle filters. In this paper we opt for the latter type of algorithms. The Monte-Carlo particle filters (PFs) have been studied a lot in the literature. The most adapted to our application context can be cast into the framework of the generic sequential importance sampling (SIS) algorithm, or SISR, when a resampling step is introduced (see, e.g., [1] or [12]). The first efficient implementation of Monte-Carlo particle filters of this type dates back to Gordon, Salmond and Smith (1993). The authors give a generic formulation of the bootstrap PF, using importance sampling ideas in the correction step and systematic use of resampling in order to avoid degeneracy of the particle filter. For this reason, the algorithm is also named SIR algorithm (sampling/importance resampling). The term interacting particle filter (IPF) is also used by [9] for the same PF algorithm. For other names of the same algorithm see [1] (page 178) and the references therein. Before describing an appropriate particle filter for our purposes let us explain how the filtering approach is feasible in our context.

5.1 The filtering approach

In our application context the initial model m is a state space model and Θ can be considered as a part of the hidden state vector. The most natural way of dealing with Θ dynamically in time is to construct copies Θ_n which remain constant in time, that is, for all n, we have $\Theta_n = \Theta_{n+1}$. Despite these constant dynamics, the version of the

particle filter that we describe in the sequel enables us to modify sequentially in time the distribution of Θ given the increasingly data vector. The corresponding augmented state vector will be denoted by $X_n^a = (X_n, \Theta_n)$. In this context, all we need to approximate is $\mathbb{E}_{\phi'}\big(\Theta_n \mid Y_{0:n} = y_{0:n}\big)$ and $\big(\mathrm{Var}_{\phi'}(\Theta_{n,i} \mid Y_{0:n} = y_{0:n})\big)_{1\leq i\leq d_{\mu_1}}$. If we have a way to approximate the intractable filtering density $p(x_n^a|y_{0:n})$ by a suitable $\hat{p}(x_n^a|y_{0:n})$, then since $x_n^a = (x_n, \theta_n)$, we can obtain an approximation of the first two conditional moments of Θ_n given the complete data vector with the help of the marginal $\hat{p}(\theta_n|y_{0:n})$. In this direction, we describe in the sequel an appropriate particle filtering technique to make this approximation.

5.2 The regularized particle filter

A particular feature of the Gaussian randomization is that the optimal variance parameters equal zero (see Theorem 1). When the variances are close to zero, and this will indeed happen after some iterations of the EM-algorithm, classical particle filters cannot perform well. This is due to the fact that the particles make a discrete approximation of the filtering distribution and consequently degenerate easier under a low variance scenario due to the resampling mechanism (this is known as sample impoverishment). A successful strategy to overcome this problem was devised in [21] with the so-called post-regularized particle filter (post-RPF). The regularization refers to the use of a kernel smoothing method to change the discrete approximation of the filtering density (induced by the weighted particles in the classical IPF) into an absolutely continuous approximation. The term post-RPF is used to indicate that the regularization step takes place after the correction step. A comprehensive exposition of RPFs and improvements can be found in [22] and some theoretical results in [15] and [16]. In [21] and [23], the authors compare the performance of the RPFs with the IPFs in some classical tracking problems. Let us now describe the adaptation of this method in our context.

We denote by K the regularization kernel (usually Epanechnikov or Gaussian) associated to X_n^a and by K_h the rescaled kernel given by:

$$K_h(x^a) = h^{-d}K(h^{-1}x^a),$$

where $h > 0$ is the bandwidth parameter and d is the dimension of the state vector. In our application, we used the Gaussian kernel and with M particles the optimal bandwidth

parameter (when the underlying density is Gaussian) is given by (see, eg., [22]):

$$h_M = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} M^{-\frac{1}{d+4}}. \tag{49}$$

Let us assume that at the end of the n-th step, the estimate of $p(x_n^a|y_{0:n})$ is given by:

$$\hat{p}(x_n^a|y_{0:n}) = \sum_{i=1}^{M} \tilde{w}_n^{(i)} K_{h_M}(x_n^a - \tilde{x}_n^{a\,(i)}),$$

where $\{(\tilde{w}_n^{(i)}, \tilde{x}_n^{a\,(i)})\}$ is the weighted sample before the regularization step. The basic filtering step can be described as follows:

- Sampling with regularization: For all $i = 1, \ldots, M$:
  - Generate $I^{(i)} \in \{1, \ldots, M\}$, with $\mathbb{P}(I^{(i)} = j) = \tilde{w}_n^{(j)}$.
  - Generate $\epsilon^{(i)} \sim K(x^a)$.
  - Compute $\tilde{x}_n^{a\,(i)} = \tilde{x}_{n^-}^{a\,(I^{(i)})} + h_M \hat{\Sigma}_n^{1/2} \epsilon^{(i)}$, where $\hat{\Sigma}_n^{1/2}$ is the square root of the empirical covariance matrix (whitening is used)
- Prediction step
  - Generate $\tilde{x}_{n+1}^{(i)} \sim p(x_{n+1}|\tilde{x}_n^{a\,(i)})$.
  - Set $\tilde{\theta}_{n+1}^{(i)} = \tilde{\theta}_n^{(i)}$.
- Correction step and regularization:
  - Set $\tilde{w}_{n+1}^{(i)} = p(y_{n+1}|\tilde{x}_{n+1}^{a\,(i)})$, for all $i = 1, \ldots, M$,
  - Regularize the weighted sample $\{(\tilde{w}_{n+1}^{(i)}, \tilde{x}_{n+1}^{a\,(i)})\}$ by taking

$$\hat{p}(x_{n+1}^a|y_{0:n+1}) = \sum_{i=1}^{M} \tilde{w}_{n+1}^{(i)} K_{h_M}(x_{n+1}^a - \tilde{x}_{n+1}^{a\,(i)}).$$

As far as the regularization is concerned, it is noteworthy to mention the recent approach presented in [24] (see also [4]). The authors proposed a particle filter where a convolution kernel is used to regularize the likelihood of the observations as well. The new filter, called convolution particle filter, seems to be more appropriate to use when the likelihood cannot be computed explicitly or when signal-to-noise ratio is very low or very high (see, eg., [24]).

5.3 A simulation case study for a plant growth model

With the purpose of testing the performance of the proposed algorithm, we performed simulations based on the true experimental conditions of one experiment conducted by the French institute for Sugar beet research in 2010 (see [2] for a detailed description of the experimental protocol). Here, we present the results of simulations in two different scenarios: (i) the complete dataset is observed, that is, 320 observations (from day 1 to day 160) and (ii) we have only access to 28 observations from this dataset, green leaf and root masses measured at the 14 dates included in $\mathcal{O}_{2010}$:

$$\mathcal{O}_{2010} = \{54, 68, 76, 83, 90, 98, 104, 110, 118, 125, 132, 139, 145, 160\},$$

which exactly correspond to the 14 dates of measurement in the 2010 experiment. In this way, we can assess the consequences of the missing information in parameter estimation.

The details of the model are given in Section 4.3. Different strategies were compared for the augmentation of the Monte Carlo sample size (geometric and quadratic) and a large constant sample size was also used to have a more ideal (but more expensive) EM algorithm (150000 particles for the restricted dataset and 40000 for the complete dataset). Two different initializations were carried out in each case and the averaging technique ([5]) was used to smooth parameter estimates after a small burn-in period. As a stopping rule we used the one proposed in [3], which claims convergence when the relative change in the estimates from three successive iterations is reasonably small [3]. A conditional version of RPF-EM is implemented with the aim of providing reasonable estimates for the noise parameters (8 conditional steps for the estimations of the noise parameters).

The parameters' real values and the estimation results are presented in Table 1. The standard errors of the variability of parameter estimates from independent runs of the algorithm (20 repetitions based on the same dataset) are also given, which corresponds to the algorithmic uncertainty. The log-likelihood values of the estimates are estimated by the means of 10 independent evaluations (the standard errors were small, approximately 0.005).

In Table 1, we give an idea of the influence of the initial distribution. Note that for the complete dataset the estimates based on these two different initializations are very close, which suggests a negligible effect of the initial distributions. However, in the case

Initialization 1							
Parameter	Real value	Initial distribution		Restricted dataset ♠		Full dataset ♣	
		Mean	Std.	Mean estimate	Std.	Mean estimate	Std.
μ_a	3.56	3.40	0.15	3.664	0.049	3.564	0.005
λ	56.6	60.0	5.0	59.73	0.22	55.68	0.12
γ_0	0.625	0.58	0.10	0.679	0.026	0.628	0.002
γ_f	0.1035	0.12	0.025	0.088	0.013	0.047	0.006
μ_γ	550.0	500.0	50.00	492.91	2.44	624.55	9.07
σ_γ	950.0	880.0	50.00	1338.75	20.25	1633.29	98.66
σ_g	0.1	0.08	-	0.106	0.001	0.108	0.001
σ_r	0.1	0.08	-	0.083	0.001	0.091	0.001
σ_q	0.05	0.035	-	0.047	0.010	0.037	0.005
$\sigma_{\gamma\gamma}$	0.05	0.035	-	0.059	0.006	0.061	0.005
Log-likelihood	-136.875 ♠ -935.350 ♣	-426.870 ♠ -17312.1 ♣	-	-135.514	0.181	-929.404	1.626

Initialization 2							
Parameter	Real value	Initial distribution		Restricted dataset ♠		Full dataset ♣	
		Mean	Std.	Mean estimate	Std.	Mean estimate	Std.
μ_a	3.56	3.60	0.15	3.668	0.058	3.566	0.003
λ	56.6	52.0	5.0	59.94	0.50	55.73	0.07
γ_0	0.625	0.7	0.10	0.686	0.036	0.628	0.002
γ_f	0.1035	0.09	0.025	0.070	0.024	0.048	0.005
μ_γ	550.0	580.0	50.00	510.24	4.60	623.14	6.69
σ_γ	950.0	1000.0	50.00	1610.93	46.88	1609.01	82.70
σ_g	0.1	0.08	-	0.106	0.002	0.108	0.001
σ_r	0.1	0.08	-	0.083	0.001	0.091	0.001
σ_q	0.05	0.035	-	0.048	0.012	0.040	0.004
$\sigma_{\gamma\gamma}$	0.05	0.035	-	0.055	0.011	0.060	0.004
Log-likelihood	-136.875 ♠ -935.350 ♣	-388.465 ♠ -4106.015 ♣	-	-135.534	0.327	-929.261	1.264

Table 1: Comparison of the estimations from two different initializations, both for the restricted (150000 particles) and the complete dataset (40000 particles). ♠: estimated log-likelihood based on the restricted dataset (14 dates); ♣: estimated log-likelihood based on the complete dataset (160 dates).

of the restricted dataset, there is a difference in the estimates of the last three parameters (γ_f, μ_γ, σ_γ) and the effect of the initialization seems to be non negligible, especially in the case of σ_γ. This can be explained by the sensitivity of model outputs Q^g and Q^r to parameters. A sensitivity analysis of the LNAS model based on Sobol's method was presented in [8] and showed that the last three parameters are the least influential ones among the six which were estimated. The increased variability of these parameters (see the corresponding standard deviations) from independent runs of the algorithm (20 repetitions) based on the same dataset is also an argument in this direction, since the great loss of information seems to affect their estimation quality. The results of log-likelihood evaluations also seem to suggest that despite that the estimated MLE could slightly differ from two different initializations, the most sensitive functional parameters can be estimated without difficulty even in the case of the restricted dataset, together with the observation and model noise parameters. By comparing the estimation results (with the same initialization) between the restricted and the complete dataset, we also conclude that the estimates of the most influential parameters are closer to the real values with the complete dataset than the restricted one, as expected, due to the great amount of missing information.

In Fig. 1, we illustrate the convergence of the estimations of the six functional parameters (for the restricted dataset) based on a single run of the algorithm. The uncertainty related to parameter estimation is evaluated by parametric bootstrap. As a better approximation of the MLE, we considered the means from the estimates of both initializations. In Table 2 we present the results that we obtained with a bootstrap sample of size 100. Except for the least influential parameters, the bootstrap means are very close to the MLE in the case of the complete dataset, but when a great amount of data is missing a small bias is introduced. By comparing the standard deviations of the estimates for the complete and the restricted dataset, it is clear that the additional amount of information reduces considerably the standard deviations of all the estimated parameters, except for the 3 least influential functional parameters. Moreover, the standard deviations of γ_f and σ_γ are prohibitively high to allow for reliable estimations. However, for all the other parameters, despite the increased standard deviations in the case of the restricted dataset, we finally get reliable estimates of the true values within the range of one standard

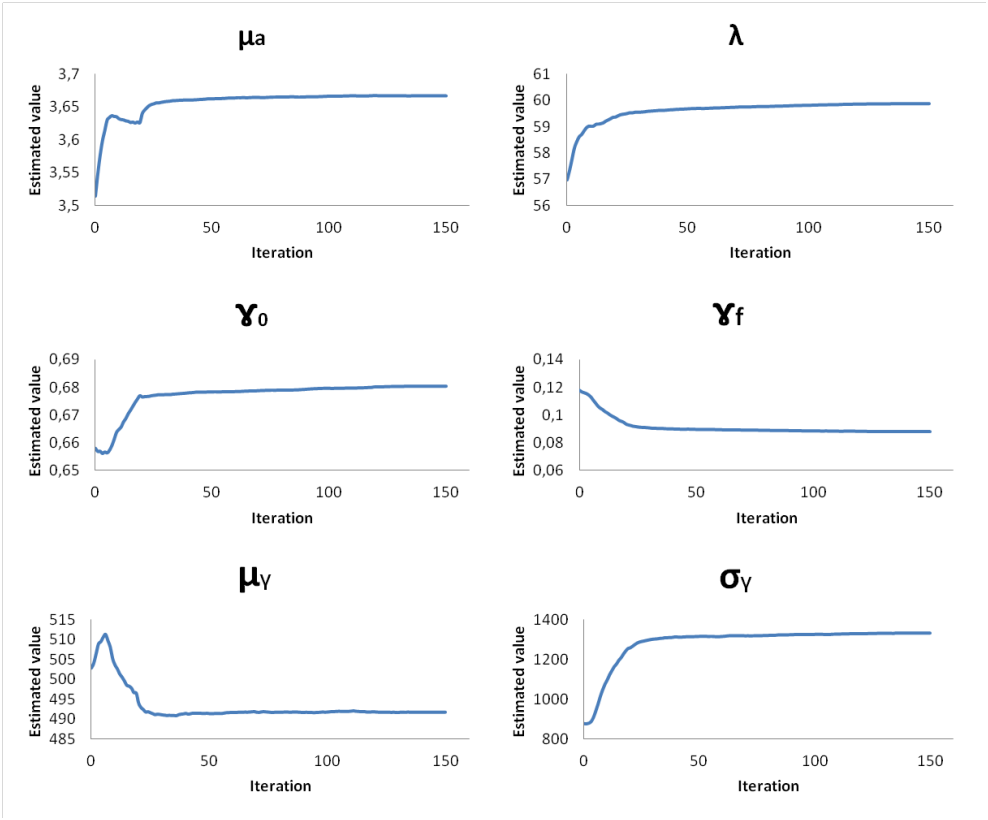

Figure 1: Evolution of the estimated values of six functional parameters: μ_a, λ, γ_0, γ_f, μ_γ and σ_γ with the RPF-EM algorithm (150000 particles) for the restricted dataset.

deviation.

In order to alleviate the burden of a large constant number of particles, we implemented and compared two different strategies of increasing the sample size during iterations, that is, a geometric and a quadratic increase. The geometric increase provided better results. In Table 3, we compare the results of the geometric increase with those that we obtained with the constant number of particles. The increase of the particles' number takes place every time that the Monte Carlo error in the evaluation of the Q-function is considered to be significant, for example, by detecting a zig-zagging in parameter estimates. Without a strong effort to optimize the increasing strategy, the computational time was reduced by 23%. The results show that the geometric increase induces slightly higher variability in the estimates among different runs of the algorithm than in the case with a constant particle number. Consequently, the gain in computational time from the decrease of the total number of simulations is counterbalanced by this effect. In this paper, we do not focus on aspects concerning the most efficient use of Monte Carlo resources and further

Parameter	Real value	Restricted dataset ♠			Full dataset ♣		
		MLE	Mean	Std.	MLE	Mean	Std.
μ_a	3.560	3.666	3.615	0.137	3.565	3.567	0.058
λ	56.6	59.8	57.7	4.6	55.7	55.7	1.2
γ_0	0.625	0.680	0.660	0.057	0.628	0.626	0.017
γ_f	0.1035	0.079	0.097	0.033	0.048	0.040	0.030
μ_γ	550.00	501.58	503.60	77.95	623.85	633.29	53.62
σ_γ	950.00	1474.84	1292.37	314.81	1621.15	1852.70	308.58
σ_g	0.1	0.106	0.101	0.016	0.108	0.106	0.005
σ_r	0.1	0.083	0.079	0.021	0.091	0.092	0.006
σ_q	0.05	0.047	0.049	0.012	0.038	0.038	0.004
$\sigma_{\gamma\gamma}$	0.05	0.057	0.054	0.016	0.061	0.052	0.006

Table 2: Uncertainty assessment in the case of both restricted and complete simulated datasets by parametric bootstrap. The means and the standard deviations are given on the basis of a 100-bootstrap sample for the restricted (14 dates) and the complete (160 dates) datasets.

investigations are needed in this direction.

Parameter	Real value	Initial distribution		Constant →		Geometric increase ↑	
		Mean	Std.	Mean estimate	Std.	Mean estimate	Std.
μ_a	3.56	3.40	0.15	3.664	0.049	3.662	0.082
λ	56.6	60.0	5.0	59.73	0.22	59.67	0.54
γ_0	0.625	0.58	0.10	0.679	0.026	0.677	0.063
γ_f	0.1035	0.12	0.025	0.088	0.001	0.089	0.001
μ_γ	550.0	500.0	50.00	492.91	2.44	495.27	7.71
σ_γ	950.0	880.0	50.00	1338.75	20.25	1372.40	32.92
σ_g	0.1	0.08	-	0.106	0.001	0.107	0.002
σ_r	0.1	0.08	-	0.083	0.001	0.085	0.001
σ_q	0.05	0.035	-	0.047	0.010	0.051	0.013
$\sigma_{\gamma\gamma}$	0.05	0.035	-	0.059	0.006	0.061	0.009
Log-likelihood	-136.875	-426.870	-	-135.514	0.181	-135.648	0.173

Table 3: Comparison of parameter estimates with two different configurations of particle numbers. →: a constant number of 150000 particles. ↑: starting from 80000, the number of particles increases geometrically until a maximum number of 150000.

6 A real-data application

In this section, we present the identification results from a real dataset with the LNAS model (described in Section 4), where parameter estimation is performed with the RPF-EM algorithm, as described in Section 5.

This dataset concerns limited experimental observations realized by the French institute for Sugar Beet research in 2010 (details of the experimental protocols are presented in [17]). Dry matter of root (denoted by Q_r) and leaves (denoted by Q_g) were collected on 50 plants at 14 different dates. The observation vector, which records at each available date the average mass per square meter, is given by Table 6.

n	54	68	76	83	90	98	104	110	118	125	132	139	145	160
$Y_{n,1}$	85.2	372.9	447.6	440.8	620.4	523.8	541.4	620.2	627.5	757.6	760.5	598.3	670.7	628.4
$Y_{n,2}$	23.1	199.8	302.4	409.2	709.2	768.1	863.9	1232.5	1498.8	1770.2	1878.2	1913.7	2118.4	2274.7

Table 4: Experimental dataset provided by the French institute for Sugar beet research (ITB) based on an experiment in 2010. 14 dates of measurement are available (denoted by n). The observation vector contains the averaged mass (from 50 plants) per square meter of green leaf compartment (denoted by $Y_{n,1}$) and root compartment (denoted by $Y_{n,2}$) in g.

The details of the model are given in Section 4.3. The estimation results are presented in Table 5 together with estimated standard errors from 20 independent runs. The Monte Carlo sample size was increased geometrically and other implementation details can be found in Section 5.3.

Note that the influence of the initialization is negligible for all the parameters but σ_γ, which is the least influential parameter. This is also suggested by the small difference of the log-likelihood evaluations at the corresponding solutions. The variability from independent runs of the algorithm was relatively high for the modeling noise parameters (σ_q, $\sigma_{\gamma\gamma}$), but by averaging a moderate number of independent runs a better precision can be obtained.

The uncertainty related to parameter estimation was assessed by parametric bootstrap, by averaging the estimates obtained from both initializations. In Table 6, we present the means and the standard deviations of a 100-bootstrap sample. Notice that the estimates of the 3 most important parameters (high sensitivity indices) are very satisfactory, with low

standard deviations relatively to their mean values. Nevertheless, the estimations from the least influential parameters are less satisfactory with higher standard deviations as expected from the synthetic case in Section 5.3. A small bias in the mean estimated values from the bootstrap sample also confirms the difficulty in estimating these parameters. Note also that the estimates of the noise parameters are satisfactory.

Parameter	Initialization I				Initialization II			
	Initial distribution		Estimation		Initial distribution		Estimation	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
μ_a	3.40	0.2	3.565	0.062	3.45	0.4	3.563	0.059
λ	58.0	3.0	59.78	0.37	57.0	5.0	59.33	0.46
γ_0	0.90	0.10	0.844	0.017	0.78	0.15	0.836	0.021
γ_f	0.18	0.05	0.190	0.002	0.20	0.05	0.198	0.001
μ_γ	650.0	40.0	642.64	5.96	630.00	50.0	641.02	6.46
σ_γ	350.0	30.00	319.59	13.62	300.0	40.0	297.78	14.80
σ_g	0.10	-	0.142	0.004	0.12	-	0.143	0.001
σ_r	0.10	-	0.164	0.004	0.12	-	0.167	0.001
σ_q	0.02	-	0.042	0.012	0.03	-	0.043	0.012
$\sigma_{\gamma\gamma}$	0.05	-	0.064	0.012	0.03	-	0.064	0.012
Log-likelihood	-173.914	-	-165.586	0.116	-169.355	-	-165.591	0.132

Table 5: Comparison of the estimations resulting from two different initializations in the case of the 2010 experimental dataset.

Parameter	MLE	Mean	Std.
μ_a	3.564	3.562	0.123
λ	59.55	59.70	3.13
γ_0	0.840	0.843	0.044
γ_f	0.194	0.182	0.053
μ_γ	642.33	654.89	62.40
σ_γ	308.69	336.62	109.95
σ_g	0.142	0.143	0.032
σ_r	0.165	0.158	0.024
σ_q	0.042	0.039	0.008
$\sigma_{\gamma\gamma}$	0.064	0.073	0.015

Table 6: Uncertainty assessment of estimates based on the 2010 experimental dataset performed by parametric bootstrap.

7 Conclusion

In the context of complex dynamic stochastic models, parameter estimation is a difficult and challenging problem. For a likelihood-based approach, many complex stochastic EM-type algorithms have been proposed to approach the maximum likelihood estimator. In real data applications the M-step is never explicit and complex numerical algorithms should be combined to make parameter estimation feasible. The proposed Gaussian randomization process allows us to make the M-step explicit and to circumvent the numerical difficulties, under the assumption that the initial model has a unique maximum likelihood estimator. The proposed RPF-EM algorithm, which consists of an E-step performed by the post-regularized particle filter circumvents classical problems such as the divergence of the filters induced by weak noises and low variance scenarios during the last iterations of the algorithm. Additionally, noise related variance parameters can easily be estimated by conditional maximization steps. Synthetic examples and results based on a real dataset suggest that the proposed algorithm performs well when confronted to scarce observations.

As a conclusion, the proposed RPF-EM algorithm based on Gaussian randomization appears as a promising alternative for robust estimation of plant growth models, and moreover in a general way for models in life sciences.

References

[1] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174 –188, feb 2002.

[2] C. Baey, A. Didier, S. Li, S. Lemaire, F. Maupas, and P.-H. Cournède. Evaluation of the predictive capacity of five plant growth models for sugar beet. In *4th international symposium on Plant Growth and Applications(PMA12), Shanghai, China*. IEEE, 2012.

[3] J. G. Booth and J. P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society Series B, Royal Statistical Society*, 61(1):265–285, 1999.

[4] F. Campillo and V. Rossi. Convolution Particle Filter for Parameter Estimation in General State-Space Models. *IEEE Transactions in Aerospace and Electronics.*, 45(3):1063–1072, 2009.

[5] O. Cappé, E. Moulines, and T. Rydén. *Inference in Hidden Markov Models.* Springer, New York, 2005.

[6] G. Celeux and J. Diebolt. The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2:73–82, 1985.

[7] Y. Chen and P.-H. Cournède. Assessment of parameter uncertainty in plant growth model identification. In *Plant Growth Modeling, Simulation, Visualization and Applications (PMA), 2012 IEEE Fourth International Symposium on*, pages 85–92, 2012.

[8] P.-H. Cournède, Y. Chen, Q. Wu, C. Baey, and B. Bayol. Development and evaluation of plant growth models: Methodology and implementation in the PYGMALION platform. *Mathematical Modelling of Natural Phenomena*, 8, 2013.

[9] P. Del Moral. Non-linear filtering: interacting particle solution. *Markov Processes Relat. Fields*, 2:555–580, 1996.

[10] B. Delyon, V. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 27:94–128, 1999.

[11] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 39:1–38, 1977.

[12] A. Doucet, N. De Freitas, and N. Gordon. *Sequential Monte Carlo methods in practice.* Springer-Verlag, New-York, 2001.

[13] C. Gaucherel, F. Campillo, L. Misson, J. Guiot, and J.J. Boreux. Parameterization of a process-based tree growth model: Comparison of optimization, mcmc and particle filtering algorithms. *Environmental Modelling and Software*, 23(10-11):1280–1288, 2008.

[14] Y. Guo, Y.T. Ma, Z.G. Zhan, B.G. Li, M. Dingkuhn, D. Luquet, and P. de Reffye. Parameter optimization and field validation of the functional-structural model greenlab for maize. *Annals of Botany*, 97:217–230, 2006.

[15] F. Le Gland, C. Musso, and N. Oudjane. An analysis of regularized interacting particle methods for nonlinear filtering. In *Proceedings of the 3rd IEEE European Workshop on Computer-Intensive Methods in Control and Signal Processing*, pages 167–174, 1998.

[16] F. Le Gland and N. Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. *Ann. Appl. Probab.*, 14(1):144–187, 2004.

[17] S. Lemaire, F. Maupas, P.-H. Cournède, J.-M. Allirand, P. de Reffye, and B. Ney. Analysis of the density effects on the source-sink dynamics in sugar-beet growth. In B.-G. Li, M. Jaeger, and Y. Guo, editors, *3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China*. IEEE Computer Society (Los Alamitos, California), November 9-12 2009.

[18] S. Lemaire, F. Maupas, P.-H. Cournède, and P. de Reffye. 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*, 2008.

[19] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons Inc., 2008.

[20] X.-L. Meng and D. B. Rubin. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika,*, 80:267–278, 1993.

[21] C. Musso and N. Oudjane. Regularization schemes for branching particle systems as a numerical solving method of the nonlinear filtering problem. In *Proceedings of the Irish Signals and Systems Conference*, Dublin, 1998.

[22] C. Musso, N. Oudjane, and F. Le Gland. Improving regularized particle filters. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods*

in Practice. New York, number 12, pages 247–271. Statistics for Engineering and Information Science, 2001.

[23] N. Oudjane and C. Musso. Multiple model particle filter. In *17ème Colloque GRETSI, Vannes 1999*, pages 681–684, 1999.

[24] V. Rossi and J.-P. Vila. Nonlinear filtering in discrete time: A particle convolution approach. *Ann. Inst. Stat. Univ. Paris*, 3:71–102, 2006.

[25] S. Trevezas and P.-H. Cournède. A sequential Monte Carlo approach for MLE in a plant growth model. *Journal of Agricultural, Biological, and Environmental Statistics*, advance online publication, doi: 10.1007/s13253-013-0134-1, 2013.

[26] S. Trevezas, S. Malefaki, and P.-H. Cournède. Simulation techniques for parameter estimation via a stochastic ECM algorithm with application to plant growth modeling. Preprint, http://hal.archives-ouvertes.fr/hal-00798695, 2013.

[27] G. Wei and M. Tanner. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, 85:699–704, 1990.

[28] P.H. Wyckoff and J.S. Clark. Predicting tree mortality from diameter growth: a comparison of maximum likelihood and bayesian approaches. *Canadian Journal of Forest Research*, 30(1):156–167, 2000.