Modeling the inter-individual variability of single-stemmed plant development

Antonin Della Della Noce, Véronique Letort, Amandine Hansart, Charlotte Baey, Gautier Viaud, Sébastien Barot, Jean-Christophe Lata, Xavier Raynaud, Paul-Henry Cournède, Jacques Gignoux

Modeling the inter-individual variability of single-stemmed plant development

Antonin Della Noce[1], Véronique Letort[1], Amandine Hansart[2], Charlotte Baey[1], Gautier Viaud[1], Sébastien Barot[3], Jean-Christophe Lata[4], Xavier Raynaud[4], Paul-Henry Cournède[1] and Jacques Gignoux[5]

[1]MICS,CentraleSupélec, University of Paris-Saclay, France
[2]CEREEP-Écotron Ile de France, France
[3]Institut de Recherche pour le Développement (IRD), France
[4]Université Pierre et Marie Curie (UPMC), France
[5]Institut d'écologie et de sciences de l'environnement (iEES), France

***Abstract*—Modeling the development of single-stemmed plants is classically done by expressing the number of phytomers in the main axis as a function of the thermal time. In many cases, a strong inter-individual variability is encountered, that cannot be straightforwardly associated to any environmental or micro-environmental factors. We propose a methodological framework to describe the heterogeneity of individual responses in a population by hierarchical or mixed-effect models : some regression parameters are random variables identically distributed in the population. The method is illustrated using experimental data collected on young *Acacia erioloba* plants, grown in CEREEP greenhouse, for which two piecewise-linear models of organogenesis are identified using Expectation-Maximization (EM) algorithm and compared according to their adequacy to the data. The E-step and the M-step of the algorithm are analytic as long as the dependence is linear with respect to the individual parameters, i.e. the latent variables of the mixed-effect model. In the general case, the posterior distribution of the individual parameters cannot be analytically determined. Monte-Carlo Markov Chains are generated with Metropolis-Hasting during the E-step to estimate the distribution. Alternatively, a Laplace approximation of the posterior distribution appeared to be quite relevant in our case. Such framework will be adapted to the identification of a complete FSPM of *Acacia* growth combining organogenesis and functioning.**

I. Introduction

The concept of thermal time and its use as time scale to model plant development has a long history since its first introduction by [1], [2]. It is usually the first and mandatory step for building a functional-structural plant model (FSPM): relying on the assumption that organogenesis is mainly monitored by thermal time [3], it consists, for single-stemmed plants (e.g. crops or trees in their early stages i.e. before the appearance of secondary axes), in expressing the number of phytomers along its main axis as a predefined function of thermal time.

Thermal time (expressed in degree-day) can be defined as:

$$\tau(t) = \int_{t_0}^{t} \max(0, T(s) - T_b)\mathrm{d}s \tag{1}$$

where t is the time expressed in days, $T(t)$ is the temperature in °C, t_0 is the emergence time of the plant, i.e. the date at which its first phytomer is observed, and T_b is the base temperature, below which the organogenesis process stops.

If the function of thermal time is chosen deterministic, plant organogenesis should be the same from an individual to another, as long as they share the same environmental conditions. However, in many cases (see Figure 1 for an illustration), a strong inter-individual variability is encountered, that cannot be straightforwardly associated to any environmental or micro-environmental factors. It has nevertheless to be accounted for in the models and cannot be neglected since it is an important characteristic that can influence the model behavior and interpretations. Besides, by integrating random effects into the model, parametric estimations at population level and at individual level are intrinsically coupled : the predicted development of an individual is consistent with the overall trend of the phenomenon in the population, with individual characteristics matching the observed population variability.

Heterogeneity was studied in a sugar beet population in [4] for the organogenesis and in [5] for a full FSPM model. The approach is based on the identification of non-linear mixed effect models with the Stochastic Approximation Expectation Maximization (SAEM) algorithm as proposed by [6]. In such framework, the parameters of the model describing the individual behavior, such as phyllochrons, are assumed to be randomly distributed over the population. In this article, we follow such a population approach using hierarchical models for plant development, developing in details each step of the identification procedure.

The method will be presented using a specific case-study that will be used all along the paper to illustrate the method and interpretation of the results. This dataset was collected on young *Acacia erioloba* plants, grown in CEREEP greenhouse. *Acacia Erioloba* is a key species of savanna ecosystems. In some areas, it is a dominant component of the tree cover, providing food (specifically nutritive pods during the dry season [7]) and shadow to many browser herbivores and possibly enriching the soil through nitrogen fixation. A noticeable plasticity between individuals was observed in [8] and explained by the plant adaptability to the multiple constraints it is subject to: fires, drought, herbivory. The design of mechanistic models to predict the dynamics of savana tree cover according to environmental conditions is of prior importance for the man-

agement of the ecosystem resources. A first attempt to capture the morphological variability induced by browser aggressions was undertaken for a closely related species, *Acacia tortillis*, growing in similar biotopes, using a GreenLab FSPM in [9]. It was able to reproduce some observed architectural patterns such as the ball-shaped crown resulting from intense grazing but was lacking proper follow-up of the individual growth trajectories to be validated, especially for the organogenesis part. This is why new experiments were conducted, with the objective of a better understanding of the growth of Acacia in controlled conditions. The study of organogenesis appears a first step for the design of a complete GreenLab FSPM, as architecture monitors biomass allocation.

Section 2 sets the notations of the problem and develops in details a step of the iterative Expectation-Maximization procedure ([10]). According to the model formulation, different versions of the EM algorithm are derived. In section 3, the method is applied to the identification of different models of organogenesis.

II. Data and Hierarchical Models

A. Experimental conditions and observed variability

A large measurement campaign has been carried out since 2013 on *Acacia erioloba* specimens in the greenhouse of CEREEP-Ecotron IleDeFrance (http://www.foljuif.ens.fr/) to analyze the dynamics of development of trees according to their environmental conditions. Plants were growing in regularly spaced pots, so that the effect of competition for light resources can be neglected. The study focuses on single axis plants, which germinated in July 2013. Weekly measurements of phytomers number, along with daily recordings of the temperature inside the greenhouse, were gathered in a dataset, totalling 567 observation points (temperatures, numbers of phytomers), corresponding to the first five months of development of 122 plants. Figure 1 gives a visualization of the number of phytomers according to the cumulated temperature, i.e. if germination date of a plant is denoted by t_0, then the associated cumulated temperature at day t is $\sum_{s=t_0}^{t} T_s$.

A global pattern can be observed on Figure 1 : the growth has essentially two stages, characterized by two distinct phyllochrons. Around this overall behavior, important variations of growth rate can be noticed from one individual to another. A piecewise linear growth can be a good approximation of the dynamics, similarly to the model adopted in [4]. The objective is to design a model that reproduces this overall dynamics and takes into account this inter-individual variability.

B. Design of two mixed-effect models

For simplicity sake, a non generic definition of mixed-effect model is given in this section. A general formulation can be found in [6].

Let $y = (y_i)_{1 \leq i \leq N}$ be the experimental observations for all individuals of the population, i.e. for all $i \in [\![1; N]\!]$, y_i is a vector in $\mathbb{R}^{n_i}$, and for all $j \in [\![1; n_i]\!]$, y_{ij} is the number of

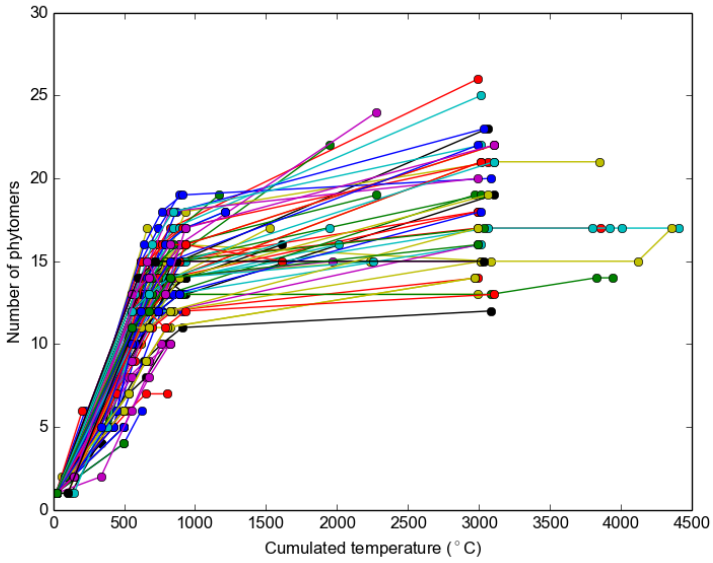

Figure 1. Number of phytomers along plant main axes with respect to the cumulated sum of temperature, i.e $\sum_{j=1}^{n} T_j$. To each colored line an individual is associated.

phytomers of plant i at the observation time t_{ij}. t_{i0} denotes the observed emergence date of plant i. Let T_b be the base temperature specific to *Acacia erioloba*, which is unknown *a priori*, its estimation being an interesting aspect of this study. Then the thermal time associated to the plant i at time $t_{ij} > t_{i0}$ is given by a discretization of equation (1).

$$\tau(t_{ij}, T_b) = \tau_{ij}(T_b) = \sum_{s=t_{i0}}^{t_{ij}-1} \max(0, T(s) - T_b) \qquad (2)$$

with the initialization $\tau(t_{i0}) = 0$. In the previous equation, $T(s)$ is the averaged temperature at day $s \in [\![t_{i0}; t_{ij}]\!]$. A model of organogenesis can therefore be formulated at the individual level as follows, for all $i \in [\![1; N]\!]$ and for all $j \in [\![1; n_i]\!]$

$$y_{ij} = 1 + f(\tau_{ij}, \phi, \alpha) + \epsilon_{ij} \qquad (3)$$

where f is a function of the thermal time, parameterized by $\phi \in \mathbb{R}^d$ and $\alpha \in \mathbb{R}^m$, and $\epsilon_{ij} \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ are additive centered Gaussian noise processes. As the germination date t_{i0} is the reference date for the thermal time, and as the number of phytomers in the plant at that date is 1, we must have $f(0, \phi, \alpha) = 0$.

To model the inter-individual variability, it can be assumed that for each plant i a parameter vector ϕ_i is sampled from a multivariate normal distribution in the population

$$\phi_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(\phi_p, \Sigma_p) \qquad (4)$$

with $(\phi_i)_{1 \leq i \leq N}$, the set of individual parameters, or random effects, $\xi = (\phi_p, \Sigma_p)$, the population parameters. α are the

fixed effects, that is to say that these parameters are constant within the population.

Conditionally to ϕ_i, the distribution of the observation model is given by equation (3)

$$y_{ij}|\phi_i \sim \mathcal{N}(1 + f(\tau_{ij}, \phi_i, \alpha), \sigma^2) \tag{5}$$

In the following, in order to alleviate the derivations, the notation $\tilde{y}_{ij} = y_{ij} - 1$ will be used. For convenience, the dependence of the model with respect to the thermal time may be occasionally written implicitly by the notation $f_i(\phi_i, \alpha) = (f(\tau_{ij}, \phi_i, \alpha))_{1 \leq j \leq n_i} \in \mathbb{R}^{n_i}$.

We have chosen two candidate models f to reproduce the organogenesis dynamics of *Acacia*. Both of them represent the organogenesis as a piecewise linear function, following the guidelines given in [4], but the first model is linear with respect to the individual parameters, whereas the other one is non-linear.

a) Model 1 : Linear with respect to the individual parameters: For simplicity sake, let us first consider the case of the linear model w.r.t. ϕ. The model is formulated as follows, for all $i \in [\![1; N]\!]$

$$f(\tau_{ij}, \phi_i, \alpha) = a_i(\tau_{ij} - \tau_0)\mathbb{I}\{\tau_{ij} > \tau_0\} + b_i(\tau_{ij} - \tau_c)\mathbb{I}\{\tau_{ij} > \tau_c\} \tag{6}$$

Here, the individual parameter is $\phi_i = (a_i, b_i)^T \in \mathbb{R}^2$ are the phyllochrons associated with the two stages of organogenesis: the first one delimited between initiation thermal time τ_0 and the thermal time of the change point τ_c and has a rate a_i ; the second one takes place after τ_c and has a rate $a_i + b_i$.

The full parameter vector for our model is denoted θ and $\theta = (\phi_p, \Sigma_p, T_b, \tau_0, \tau_c, \sigma^2)$. The constant parameters that are connected to the notion of thermal time are called the *thermal* parameters and are denoted $\alpha = (T_b, \tau_0, \tau_c)$. To underline the linear aspect with respect to the individual parameters, we will use the notation $f_i(\phi_i, \alpha) = f_i(\alpha)\phi_i$. A descriptive model which is linear with respect to the random effects and non-linear with respect to the fixed effects was studied in the case of orange tree growth in [11].

b) Model 2 : non-linear mixed effects: The assumptions of the previous model are relaxed concerning the thermal times. The equation is the same as in (6), except that here τ_0 and τ_c are also considered variable in the population and $\phi_i = (a_i, b_i, \tau_0^i, \tau_c^i)^T \in \mathbb{R}^4$. The full parameter vector is now $\theta = (\phi_p, \Sigma_p, T_b, \sigma^2)$ with $\phi_p \in \mathbb{R}^4$, and the only fixed effect parameter being the base temperature $\alpha = T_b$.

In a frequentist approach, we want to estimate the parameters θ by maximizing the likelihood given the observations. The likelihood is given by:

$$p_\theta(\tilde{y}) = \prod_{i=1}^{N} \int_{\mathbb{R}^d} \mathcal{N}(\phi_i; \phi_p, \Sigma_p) \left[\prod_{j=1}^{n_i} \mathcal{N}(\tilde{y}_{ij}; f(\tau_{ij}, \phi_i, \alpha), \sigma^2) \right] \mathrm{d}\phi_i$$

where we denote $\mathcal{N}(x; \mu, \Sigma)$ the probability density function of $\mathcal{N}(\mu, \Sigma)$ evaluated at x. The evaluation and/or the maximization of such likelihood, in the case of both model 1 and model 2, is analytically intractable. Expectation-Maximization (EM) algorithms, designed for the estimation in incomplete data problems, is the intrinsic method to identify hierarchical models, by generating a sequence (θ^n) such that $(p_{\theta^n}(y))$ is increasing. The next subsection specifies how such algorithm is applied to the identification of our two models.

III. An iteration of the EM algorithm

We recall the principles of the EM algorithm from [10], and how it can be adapted to different model formulations. Let us suppose that we have an estimate θ^n of the parameter θ from the initialization or from a previous iteration. The observation log-likelihood associated to individual i $\ell_i(\theta) = \log p_\theta(\tilde{y}_i)$ can then be decomposed as follows

$$\begin{aligned}
\ell_i(\theta) &= \int_{\mathbb{R}^d} \ell_i(\theta) p_{\theta^n}(\phi_i|\tilde{y}_i) \mathrm{d}\phi_i && (7)\\
&= \int_{\mathbb{R}^d} \log p_\theta(\tilde{y}_i, \phi_i) p_{\theta^n}(\phi_i|\tilde{y}_i) \mathrm{d}\phi_i && (8)\\
&\quad - \int_{\mathbb{R}^d} \log p_\theta(\phi_i|\tilde{y}_i) p_{\theta^n}(\phi_i|\tilde{y}_i) \mathrm{d}\phi_i && (9)\\
&= Q_i(\theta|\theta^n) + \mathrm{H}_i(\theta|\theta^n) && (10)
\end{aligned}$$

where the derivation of the lines (8) and (9) is obtained by applying Bayes' rule. The first term (8) is the expected joint log-likelihood $Q_i(\theta|\theta^n) = \mathbb{E}_{\theta^n}[\log p_\theta(\tilde{y}_i, \phi_i)|\tilde{y}_i]$. The second term (9) $\mathrm{H}_i(\theta|\theta^n) = -\mathbb{E}_{\theta^n}[\log p_\theta(\phi_i|\tilde{y}_i)|y_i]$ is the cross entropy of the two posterior distributions. Using Jensen inequality, it can be proved that for all θ, $\mathrm{H}_i(\theta|\theta^n) \geq \mathrm{H}_i(\theta^n|\theta^n)$. To maximize $\ell_i(\theta)$, we are therefore interested in maximizing the total expected joint likelihood

$$\theta^{n+1} = \operatorname*{argmax}_\theta \sum_{i=1}^{N} Q_i(\theta|\theta^n) = \operatorname*{argmax}_\theta \; Q(\theta|\theta^n) \tag{11}$$

A qualitative description of EM can be formulated as follows : at each iteration of the algorithm, we use our current estimate θ^n to *generate* the latent variables ϕ_i, then we maximize the joint log-likelihood $Q(\theta|\theta^n)$.

The next paragraphs specify the expectation and the maximization steps for models 1 and 2.

A. Expectation step

The objective of this step is to find a close form of $Q(\theta|\theta^n)$. For all $i \in [\![1; N]\!]$, let ϕ_i be a realization of an individual parameter, then for all θ

$$\log p_\theta(y_i, \phi_i) = \log p_\theta(\tilde{y}_i|\phi_i) + \log p_\theta(\phi_i)$$

$$\log p_\theta(\tilde{y}_i, \phi_i) = -\frac{n_i + d}{2} \log(2\pi) - n_i \log(\sigma)$$

$$- \frac{\|\tilde{y}_i - f_i(\phi_i, \alpha)\|^2}{\sigma^2} - \log \det(\Sigma_p) - \frac{1}{2}(\phi_i - \phi_p)^T \Sigma_p^{-1} (\phi_i - \phi_p)$$

The expectation step therefore requires the evaluation of the following functions

$$\begin{aligned} \mathrm{MSE}_i^n(\alpha) &= \mathbb{E}_{\theta^n}\left[\|\tilde{y}_i - f_i(\phi_i,\alpha)\|^2|\tilde{y}_i\right] \quad (12)\\ D_i^n(\phi_p,\Sigma_p) &= \mathbb{E}_{\theta^n}\left[(\phi_i-\phi_p)^T\Sigma_p^{-1}(\phi_i-\phi_p)|\tilde{y}_i\right] \quad (13)\end{aligned}$$

a) Model 1: For model 1, thanks to the linearity with respect to ϕ_i, the posterior distribution $p_{\theta^n}(\phi_i|y_i)$ can be analytically derived, giving close form expressions for the previous functions with respect to θ. Let us derive those expressions. $p_{\theta^n}(\phi_i|y_i)$ is proportional to the following distribution

$$\begin{aligned} p_{\theta^n}(\phi_i|y_i) &\propto p_{\theta^n}(y_i|\phi_i)p_{\theta^n}(\phi_i) \quad (14)\\ &\propto \exp\left(-\frac{\|\tilde{y}_i - f_i(\alpha^n)\phi_i\|^2}{2(\sigma^n)^2}\right. \quad (15)\\ &\left. -\frac{1}{2}(\phi_i-\phi_p^n)^T(\Sigma_p^n)^{-1}(\phi_i-\phi_p^n)\right) \quad (16)\end{aligned}$$

The previous equation shows that $\phi_i|y_i$ is normally distributed, and its covariance matrix and mean are respectively given by

$$\begin{aligned} \Sigma_i^n &= \left((\Sigma_p^n)^{-1} + \frac{f_i(\alpha^n)^T f_i(\alpha^n)}{(\sigma^n)^2}\right)^{-1} \quad (17)\\ \mu_i^n &= \Sigma_i^n\left((\Sigma_p^n)^{-1}\phi_p^n + \frac{f_i(\alpha^n)^T\tilde{y}_i}{(\sigma^n)^2}\right) \quad (18)\end{aligned}$$

The expectation (13) is therefore

$$\begin{aligned} D_i^n(\phi_p,\Sigma_p) &= \mathbb{E}_{\theta^n}\left[(\phi_i-\mu_i^n)^T\Sigma_p^{-1}(\phi_i-\mu_i^n)|\tilde{y}_i\right]\\ &\quad + (\mu_i^n-\phi_p)^T\Sigma_p^{-1}(\mu_i^n-\phi_p)\\ D_i^n(\phi_p,\Sigma_p) &= \mathrm{Tr}\left(\Sigma_p^{-1}\Sigma_i^n\right) + (\mu_i^n-\phi_p)^T\Sigma_p^{-1}(\mu_i^n-\phi_p) \quad (19)\end{aligned}$$

Similarly, for expectation (12)

$$\mathrm{MSE}_i^n(\alpha) = \mathrm{Tr}\left(f_i(\alpha)^T f_i(\alpha)\Sigma_i^n\right) + \|\tilde{y}_i - f_i(\alpha)\mu_i^n\|^2 \quad (20)$$

The expected joint log-likelihood is obtained by combining all the individual terms

$$\begin{aligned} Q(\theta|\theta^n) &= -\frac{M+Nd}{2}\log(2\pi)\\ &- M\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^N \mathrm{MSE}_i^n(\alpha) \quad (21)\\ &- \frac{N}{2}\log\det(\Sigma_p) - \frac{1}{2}\sum_{i=1}^N D_i^n(\phi_p,\Sigma_p)\end{aligned}$$

where the total number of observations is $M = \sum_{i=1}^N n_i$.

b) Model 2: : The situation in the non-linear case is more complex since the first and second order moments of $p_{\theta^n}(\phi_i|y_i)$ cannot be derived analytically as in the linear case. The idea is then to estimate the quantities of interest MSE and D_i^n by generating samples from the distribution. It is only known up to a multiplicative constant

$$\begin{aligned} p_{\theta^n}(\phi|\tilde{y}_i) &\propto \exp\left(-\frac{\|\tilde{y}_i - f_i(\phi_i,T_b^n)\|^2}{2(\sigma^n)^2}\right.\\ &\left. -\frac{1}{2}(\phi_i-\phi_p^n)^T(\Sigma_p^n)^{-1}(\phi_i-\phi_p^n)\right) \quad (22)\\ p_{\theta^n}(\phi|\tilde{y}_i) &\propto \exp(-J_i^n(\phi_i)) \quad (23)\end{aligned}$$

MCMC techniques, such as Metropolis-Hasting, enables the generation of a Markov Chain, whose stationary distribution is the target distribution (22), and the moments of interest can be computed by the ergodic theorem. A MCMC simulation within the Expectation step of $p_{\theta^n}(\phi_i|\tilde{y}_i)$ is therefore adopted here (for a a more detailed description on MCMC-EM methods, see [12]).

An adaptative Metropolis-Hasting algorithm, introduced by [13], is used in this paper. The procedure is described in the following frame

Algorithm 1 Adaptative Metropolis-Hasting algorithm[14]

Let L be the length of the chain, $\lambda_0 > 0$, $\beta^* \in]0,1[$ and $(\gamma_k)_{k\in\mathbb{N}}$ a sequence of learning rates.

for $i \in [\![1;N]\!]$ **do**

Initialization

find *good* values for the starting point of the chain ϕ_i^0, the initial covariance of the proposal distribution Σ_0, and the initial mean μ_0 for covariance update

Iterations

for $k \in [\![1;L]\!]$ **do**

1. sample a new candidate
$z \sim \mathcal{N}(\phi_{k-1}, \lambda_{k-1}\Sigma_{k-1})$
2. select z with probability
$\beta(\phi_i^{k-1}, z) = \min\left(\frac{p_{\theta^n}(z|\tilde{y}_i)}{p_{\theta^n}(\phi_i^{k-1}|\tilde{y}_i)}, 1\right)$
3. select z with probability β : $\phi_i^k = z$
or reject it $\phi_i^k = \phi_i^{k-1}$.
4. update the mean
$\mu_k = \mu_{k-1} + \gamma_k(\phi_i^k - \mu_{k-1})$
5. update the covariance
$\Sigma_k = \Sigma_{k-1} + \gamma_k\left((\phi_i^k-\mu_k)(\phi_i^k-\mu_k)^T - \Sigma_{k-1}\right)$
6. Update the scale of the covariance
$\lambda_k = \exp\left(\log(\lambda_{k-1}) + \gamma_k(\beta(\phi_i^{k-1},z) - \beta^*)\right)$

end for

Add the chain (ϕ_i^k) to the stack

end for

return the stack of all chains $(\phi_i^k)_{1\le i\le N}^{1\le k\le L}$

Efficiency of the MCMC algorithm is sensitive to its initialization. We propose to choose ϕ_i^0 maximizing $p_{\theta^n}(\phi|\tilde{y}_i)$, which is equivalent to minimizing the quantity:

$$J_i^n(\phi_i) = \frac{1}{(\sigma^n)^2}\|\tilde{y}_i - f_i(\phi_i,T_b^n)\|^2 + (\phi_i-\phi_p^n)^T(\Sigma_p^n)^{-1}(\phi_i-\phi_p^n) \quad (24)$$

As the function f_i is non-smooth, we have chosen to use the Nelder-Mead algorithm, embedded in the function `optimize` in the `Optim` package (JULIA language), to compute ϕ_i^0. At this point, the posterior density has high values, and the acceptance rate α is likely to be non zero (unless this maximum is too sharp). Therefore, the generated chain will potentially not get stuck in an area where the target density is low. The covariance of the proposal is initialized at the population covariance, so that the update of the parameter is adapted to the scale of each component ϕ_i.

We use the generated samples $(\phi_i^k)^{1\leq k\leq N}$ for the Monte-Carlo estimation of MSE_i^n and D_i^n for each individual i

$$\text{MSE}_i^n(T_b) \approx \frac{1}{L-K+1}\sum_{k=K}^{L}\|\tilde{y}_i - f_i(\phi_i^k, T_b)\|^2$$
$$D_i^n(\phi_p, \Sigma_p) \approx \text{Tr}(\Sigma_p^{-1}\hat{\Sigma}_i^n) + (\hat{\mu}_i^n - \phi_p)^T\Sigma_p^{-1}(\hat{\mu}_i^n - \phi_p)$$

where K is a burn-in period, $\hat{\mu}_i^n = \frac{1}{L-K+1}\sum_{k=K}^{L}\phi_i^k$ is the empirical mean and $\hat{\Sigma}_i^n = \frac{1}{L-K+1}\sum_{k=K}^{L}(\phi_i^k - \hat{\mu}_i^n)(\phi_i^k - \hat{\mu}_i^n)^T$ is the empirical covariance. Injecting those expressions in 21, we obtain the joint-likelihood of the expectation step.

c) Alternative E-step for Model 2: An alternative to MCMC simulation of the posterior distribution appears when considering its shape. As it is shown in the Results section, $p_{\theta^n}(\phi|\tilde{y}_i)$ are unimodal distributions, very sharp around their respective modes. Such situation motivated [15] and [16] to approximate the target by a multivariate Gaussian, either by a linearization in the neighborhood of the fixed effects, either by using the Laplace approximation.

The qualitative principle of such approximation is as follows: near the argmax of $p_{\theta^n}(\phi_i|y_i)$ (that is to say near the argmin of J_i^n), denoted by μ_i^n, the distribution can be approximated by the second order development

$$J_i^n(\phi_i) = J_i^n(\mu_i^n) + \frac{1}{2}(\phi_i - \mu_i^n)^T\partial^2_{\phi_i}J_i^n(\mu_i^n)(\phi_i - \mu_i^n) \quad (25)$$

as the gradient is zero, leading to

$$p_{\theta^n}(\phi_i|\tilde{y}_i) \propto \exp\left(\frac{1}{2}(\phi_i - \mu_i^n)^T\partial^2_{\phi_i}J_i^n(\mu_i^n)(\phi_i - \mu_i^n)\right) \quad (26)$$

If $\bar{\phi}_i$ is an isolated maximum, the hessian $\partial^2_{\phi_i}J_i^n(\bar{\phi}_i)$ is positive definite, and $p_{\theta^n}(\phi_i|\tilde{y}_i)$ can be approximated by a Gaussian distribution $\mathcal{N}\left(\mu_i^n, -\left(\partial^2_{\phi_i}J_i^n(\mu_i^n)\right)^{-1}\right)$.

A problem specific to model 2 is the fact that it is non-smooth (because of the indicator function) and taking second order derivatives of $f_i(\phi_i, T_b)$ might be numerically unstable. A $\mathcal{C}^\infty$ approximation of the model 2 was considered, using a sharp sigmoid function σ_ϵ as a surrogate of the indicator function $\mathbb{I}\{.>0\}$[1].

$$\sigma_\epsilon(x) = \frac{1}{1+\epsilon\exp\left(-\frac{x}{\epsilon}\right)} \xrightarrow[\epsilon\to 0]{L^2(\mathbb{R})} \mathbb{I}\{x>0\} \quad (27)$$

Similar differentiable approximations of piecewise-model were used in [17]. This enables to have analytic expressions of the derivative in every points of the parameter space.

The minimum μ_i^n of J_i^n can be obtained numerically, once again by using Nelder-Mead or Gauss-Newton algorithms for instance. The expression of the covariance matrix Σ_i^n as a function of the model derivatives is

$$\Sigma_i^n = \left((\Sigma_p^n)^{-1} + \frac{1}{(\sigma^n)^2}\sum_{j=1}^{n_i}\partial_{\phi_i}f(\tau_{ij},\mu_i^n)\partial_{\phi_i}f(\tau_{ij},\mu_i^n)^T \right. \quad (28)$$
$$\left. + \frac{1}{(\sigma^n)^2}\sum_{j=1}^{n_i}(f(\tau_{ij},\mu_i^n) - \tilde{y}_{ij})\partial^2_{\phi_i}f(\tau_{ij},\mu_i^n)\right)^{-1} \quad (29)$$

The previous equation is consistently equal to (17) if f is linear with respect to the individual parameters. The positivity of the matrix holds since μ_i^n is a minimizer of J_i^n.

The deviation $D_i^n(\phi_p, \Sigma_p)$ has with these notations exactly the same expression as in (19). However, because of the non-linearity, the mean square error MSE_i^n has to be estimated by an empirical mean, using an i.i.d sample $(\phi_i^k)_{1\leq k\leq L} \sim \mathcal{N}(\mu_i^n, \Sigma_i^n)$

$$\text{MSE}_i^n(T_b) \approx \frac{1}{L}\sum_{k=1}^{L}\|\tilde{y}_i - f_i(\phi_i^k, T_b)\|^2 \quad (30)$$

B. Maximization step

This section derives the method to find the maximum argument of the function $Q(\theta|\theta^n)$ for model 1 and model 2.

a) Model 1: In the case of the linear model, the update of the parameters ϕ_p, Σ_p, σ is analytic, while the update of the fixed effects α (base temperature T_b and thermal times τ_0, τ_c) are performed by the numerical minimization of the functional $\text{MSE}^n = \sum_{i=1}^{N}\text{MSE}_i^n(\alpha)$.Let us start by the updates of ϕ_p and Σ_p which are obtained by solving the respective equations $\partial_{\phi_p}Q = 0$ and $\partial_{\Sigma_p}Q = 0$.

$$\phi_p^{n+1} = \frac{1}{N}\sum_{i=1}^{N}\mu_i^n \quad (31)$$
$$\Sigma_p^{n+1} = \frac{1}{N}\sum_{i=1}^{N}\Sigma_i^n + (\mu_i^n - \phi_p^{n+1})(\mu_i^n - \phi_p^{n+1})^T \quad (32)$$

[1]This $\mathcal{C}^\infty$ approximation of the indicator function is obtained by taking its convolution with the approximate Dirac distribution $\frac{1}{\epsilon}\sigma_\epsilon(x)(1-\sigma_\epsilon(x)) \underset{\epsilon\to 0}{\to} \delta_0$ in the sense of the distributions.

As for the fixed effects $\tilde{\alpha}$, it is interesting to notice that the quantity to minimize can be written as

$$\text{MSE}^n(\alpha) = \sum_{i=1}^{N} \|\tilde{y}_i - f_i(\alpha)\mu_i^n\|^2 + \sum_{i=1}^{N} \text{Tr}(f_i(\alpha)^T f_i(\alpha)\Sigma_i^n) \tag{33}$$

which appears as a bias and variance decomposition. The update of the fixed -effect estimate is obtained by minimizing this mean square error term.

$$\alpha^{n+1} = \underset{\alpha}{\text{argmin}}\ \text{MSE}^n(\alpha) \tag{34}$$

This minimization is performed using a simplex algorithm as for equation 24. This procedure estimates at the same time the optimal change points τ_0 and τ_c, and also the base temperature T_b. As a matter of fact, the procedure does not take into account the fact the model is piecewise-linear. A methodology specific to the identification of change points in piecewise-linear models were derived in [18]. Unfortunately, such procedure cannot be straightforwardly applied here as it requires to estimate the slopes and the change points at the same time, whereas in our case the slopes are computed during the expectation stage, prior to the estimation of the change points. Finally the update of the observation noise parameter σ is given by the equation $\partial_\sigma Q = 0$, which leads to

$$\sigma^{n+1} = \sqrt{\frac{\text{MSE}^n(\alpha^{n+1})}{M}} \tag{35}$$

b) Model 2: The maximization step in the case of the non-linear model follows exactly the same pattern as previously, except that the characteristics of the posterior distribution μ_i^n and Σ_i^n have to be replaced by their Monte-Carlo estimates $\hat{\mu}_i^n$ and $\hat{\Sigma}_i^n$ or by the characteristics of the Gaussian posterior derived by Laplace approximation. Besides, the MSE function is only dependent on the base temperature T_b, such function can be efficiently minimized using univariate optimization methods like the Brent method.

IV. Results

Parameter identification of hierarchical models enables multi-level predictions : at the population level, the distribution of the observations depends on the estimated parameters ϕ_p, Σ_p and on σ^2; at the individual level, a distribution of observations is also identified through the individual estimates.

A. Identification of Model 1

In the case of the linear model, the EM algorithm was run for 200 iterations, starting from initialization values that are presented in Table IV-A. After 150 iterations, the sequence of $(Q(\theta^{n+1}|\theta^n))_n$ was quite stable, so as the standard deviations of the observation $(\sigma^n)_n$. The increase of $Q(\theta^{n+1}|\theta^n)$ ensures that the parameters update is going in a good direction, just like the decrease of the standard deviation. The following table gives the estimates of the model parameters after convergence.

Parameter	Initial value	Estimate
base temperature T_b	10	14.5°C
initiation thermal time τ_0	10	32.2 degree day
thermal time of the change point τ_c	100	208 degree day
observation noise σ	1.0	0.697
first phyllochron $\frac{1}{\phi_p^1}$	50	14.6 degree day
second phyllochron $\frac{1}{\phi_p^1 + \phi_p^2}$	100	49.6 degree day
variability of phyllochron 1	± 1	$\pm$ 4.6 degree day
variability of phyllochron 2	± 1	$\pm$ 13 degree day
correlation(ϕ_p^1, ϕ_p^2)	0	0.709

Table I
PARAMETERS ESTIMATES OF MODEL 1

At the population level (Figure 2), the observations made at thermal time $y(\tau) = 1 + f(\tau)^T\phi + \epsilon$ are normally distributed:

$$y(\tau) \sim \mathcal{N}\left(1 + f(\tau)^T\phi_p, \sigma^2 + \text{Tr}\left(f(\tau)f(\tau)^T\Sigma_p\right)\right) \tag{36}$$

which gives the confidence intervals of the prediction distribution.

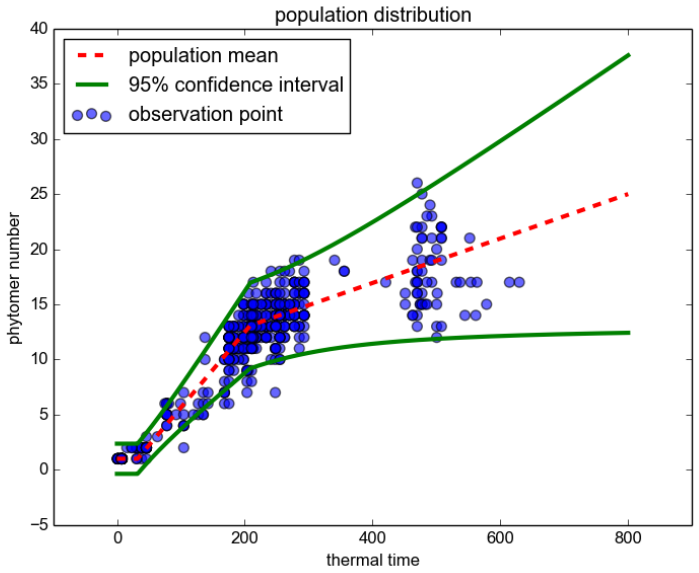

Figure 2. Predicted distribution of the observations at the population level

We can also check in Figure 3 whether the shape of the empirical distribution of the individual means $(\mu_i^n)_{1\leq i\leq N}$ is close to a Gaussian. If the number of individuals in the population is too small, this graphical representation might be inappropriate.

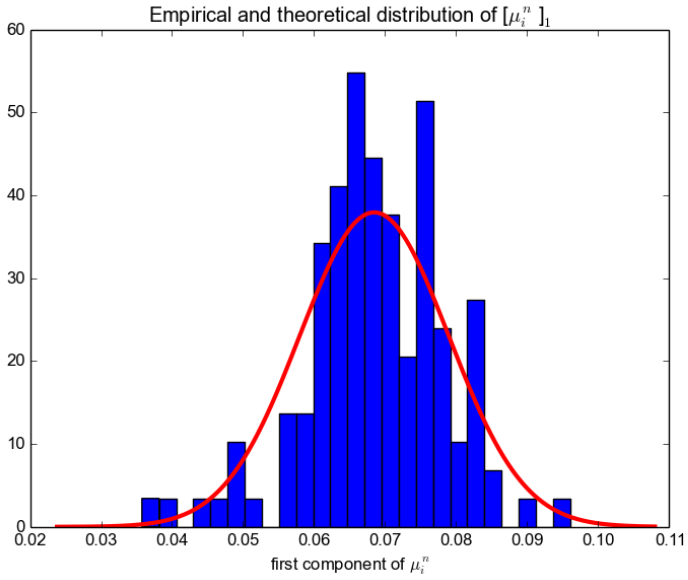

Figure 3. Empirical and theoretical distributions of the individual means

At the individual level, the predicted distribution of observations y_i for some individual i can be represented as in Figure 4, replacing ϕ_p and Σ_p by the characteristics of the posterior distribution, μ_i^n and Σ_i^n in (36).

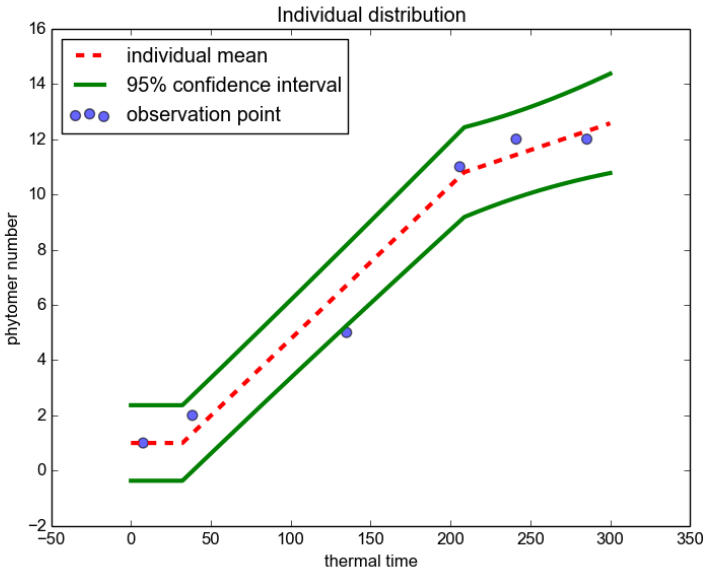

Figure 4. Prediction of the observations distribution at the individual level

In conclusion, the linear model shows a good adequacy with the observations. As model 2 is a more generic formulation of model 1, in the sense that one can go from one to another by setting the variance of τ_0 and τ_c to zero, the parameter values estimated in the linear case can be used to initialize the EM algorithm performed in the non-linear case.

B. Identification of Model 2

For the non-linear model, a MCMC-EM was first performed for 100 iterations starting from the values estimated in the previous model. The length of the Monte-Carlo Markov chain was set to 3000, with a burn-in period of 1000. The parameter λ of Metropolis-Hasting algorithm was first initialized to 1 and the optimal acceptance rate was set to $\beta^* = 0.234$, following advice in [13]. Such model did not improve significantly the mean squared error compared to the first one, in terms of mean square error and observation noise: the initial value of σ was 0.697 (see previous model) and it ends up being equal to 0.67. The base temperature was also very stable, the final estimate being 14.4. Nevertheless it was observed, by considering the sample generated by Metropolis-Hasting algorithm, that the Gaussian approximation could be a quite good estimate of the posterior distribution in itself, leading us to implement a Laplace approximation method.

After convergence of the EM with the Laplace approximation, we have noticed that the graph of the predicted population distributions and individual distributions were quite similar to the ones obtained in the linear case (Model 1). The standard deviations of parameters τ_0 and τ_c were found to be critically low, respectively 9×10^{-3} $°C.d$ and 1×10^{-2} $°C.d$. At this point, there are two possible explanations of such results :

- either modeling initiation thermal time and thermal time of the change point as random effects is inappropriate and the linear assumption is justified
- or the fact to initialize the EM algorithm in the region of parameter space corresponding to the configuration where those two parameters were constant at the population level *maintains* them in this disposition. In that case, a new initialization procedure needs to be investigated.

V. Discussion and Perspectives

The organogenesis of young *Acacia erioloba* provides a new and interesting case of application of hierarchical models. Such population approach seems relevant: the heterogeneity within the population is quite well represented by simple linear models, whose identification with EM algorithm is analytic. The non-linear version brought only minor improvements, due to the fact that the thermal times of change point and initiation could be appropriately modeled by fixed effects in our case.

The methodology developed could be easily extended to other non-linear mixed effects models. Indeed, the first method only requires linearity with respect to the individual parameters, and such assumption can be justified in a wide range of dynamics. In our case, the model is a linear combination of piecewise linear functions, but other families of non-linear functions, such as sinusoids, Gaussian kernels would have lead to the same derivations. Identifying a linearized version of the model provides relevant initialization values to EM algorithms applied on a generic form of the model, with non-linearity with respect to the individual parameters.

Although the presentation and illustration of our methodology was the core of this paper, this study is also a step towards

a more complete modeling of *Acacia erioloba* development using a complete GreenLab FSPM. For further stages of development, with branchings and reiterations, modeling the rhythm of appearance of phytomers as a function of the thermal time is no longer sufficient as current architecture plays a role. The development of the branching patterns involves complex processes that can be modeled through the retroaction of plant functioning on architecture (see [19] within the GreenLab framework) or with stochasticity ([20]).

Finally, the real objective is to derive a full population model for *Acacia erioloba* based on an individual FSPM in order to account for the inter-individual variability. Likewise, and as introduced by [5], both organogenesis and functional parameters should be modelled as mixed effects. The main difficulty will rely on the joint estimation of organogenesis and functioning, especially in the case of interaction between both.

REFERENCES

[1] W. C. Allee, O. Park, A. E. Emerson, T. Park, K. P. Schmidt *et al.*, *Principles of animal ecology.* WB Saundere Co. Ltd., 1949, no. Edn 1.

[2] D. Trudgill, A. HONEK, D. Li, and N. Straalen, "Thermal time–concepts and utility," *Annals of Applied Biology*, vol. 146, no. 1, pp. 1–14, 2005.

[3] G. G. Gramig and D. E. Stoltenberg, "Leaf Appearance Base Temperature and Phyllochron for Common Grass and Broadleaf Weed Species," *Weed Technology*, vol. 21, no. 1, pp. 249–254, 2007.

[4] C. Baey, A. Didier, S. Lemaire, F. Maupas, and P.-H. Cournède, "Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model," *Ecological Modelling*, vol. 263, pp. 56–63, aug 2013.

[5] C. Baey, S. Trevezas, and P.-H. Cournède, "A non linear mixed effects model of plant growth and estimation via stochastic variants of the em algorithm," *Communications in Statistics-Theory and Methods*, vol. 45, no. 6, pp. 1643–1669, 2016.

[6] E. Kuhn and M. Lavielle, "Maximum likelihood estimation in nonlinear mixed effects models," *Computational Statistics and Data Analysis*, vol. 49, no. 4, pp. 1020–1038, 2005.

[7] M. E. Barnes, "Seed predation, germination and seedling establishment of Acacia erioloba in northern Botswana," *Journal of Arid Environments*, vol. 49, pp. 541–554, 2001. [Online]. Available: http://www.idealibrary.com

[8] C. L. Seymour, "Grass, rainfall and herbivores as determinants of Acacia erioloba (Meyer) recruitment in an African savanna," *Plant Ecology*, vol. 197, no. 1, pp. 131–138, 2008.

[9] A. Mathieu, V. Letort, R. Gosset, J. Gignoux, and P. H. Cournède, "Simulation of morphological plasticity of Acacia tortilis in response to herbivore attacks," in *Plant Growth Modeling, Simulation, Visualization and Applications, Proceedings - PMA09*, 2010, pp. 115–118.

[10] 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*, vol. 39, no. 1, pp. 1–38, 1977. [Online]. Available: http://www.jstor.org/stable/2984875

[11] J. Pinheiro and D. Bates, *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, 2006.

[12] G. Fort and E. Moulines, "Convergence of the monte carlo expectation maximization for curved exponential families," *Annals of Statistics*, pp. 1220–1259, 2003.

[13] H. Haario, E. Saksman, and J. Tamminen, "Adaptive proposal distribution for random walk Metropolis algorithm," *Computational Statistics*, vol. 14, no. 3, p. 375, 1999.

[14] Y. Chen, "Inférence bayésienne dans les modèles de croissance de plantes pour la prévision et la caractérisation des incertitudes," Ph.D. dissertation, nov 2014. [Online]. Available: http://arxiv.org/abs/1011.1669 http://dx.doi.org/10.1088/1751-8113/44/8/085201

[15] M. J. Lindstrom, D. M. Bates, and M. J. Lindstrom, "Nonlinear Mixed Effects Models for Repeated Measures Data," *Biometrics*, vol. 46, no. 3, pp. 673–687, 1990.

[16] E. Vonesh, "A note on the use of Laplace's approximation for nonlinear mixed-effects models," *Biometrika*, vol. 83, no. 2, pp. 447–452, 1996. [Online]. Available: http://biomet.oupjournals.org/cgi/doi/10.1093/biomet/83.2.447

[17] A. Tishler and I. Zang, "A Maximum Likelihood Method for Piecewise Regression Models with a Continuous Dependent Variable," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 30, no. 2, pp. 116–124, 1981. [Online]. Available: http://www.jstor.org/stable/2346380

[18] D. M. Hawkins, "Point Estimation of the Parameters of Piecewise Regression Models," *Applied Statistics*, vol. 25, no. 1, p. 51, 1976. [Online]. Available: http://www.jstor.org/stable/2346519?origin=crossref

[19] A. Mathieu, P. H. Cournède, V. Letort, D. Barthélémy, and P. De Reffye, "A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition," *Annals of Botany*, vol. 103, no. 8, pp. 1173–1186, 2009.

[20] C. Loi and P.-H. Cournède, "Generating functions of stochastic L-systems and application to models of plant development," in *DMTCS proceedings*, 2008, pp. 325–338.