HAL

archives-ouvertes.fr

Markov and semi-Markov switching linear mixed models for identifying forest tree growth components.

Florence Chaubert-Pereira, Yann Guédon, Christian Lavergne, Catherine Trottier

▶ To cite this version:

INRIA

Markov and semi-Markov switching linear mixed models for identifying forest tree growth components

Florence Chaubert-Pereira — Yann Guédon — Christian Lavergne — Catherine Trottier

N° 6618

August 19, 2008

Thème BIO

Rapport de recherche

Markov and semi-Markov switching linear mixed models for identifying forest tree growth components

Florence Chaubert-Pereira[*][†] , Yann Guédon[*][†], Christian Lavergne[‡] , Catherine Trottier[‡]

Abstract: Observed tree growth is the result of three components: (i) an endogenous component which is assumed to be structured as a succession of roughly stationary phases separated by marked change points asynchronous between individuals, (ii) a time-varying environmental component which is assumed to take the form of local fluctuations synchronous between individuals, (iii) an individual component which corresponds to the local environmental of each tree. In order to identify and to characterize these three components, we propose to use semi-Markov switching linear mixed models, i.e. models that combine linear mixed models in a semi-markovian manner. The underlying semi-Markov chain represents the succession of growth phases (endogenous component) while the linear mixed models attached to each state of the underlying semi-Markov chain represent in the corresponding growth phase both the influence of time-varying environmental covariates (environmental component) as fixed effects and inter-individual heterogeneity (individual component) as random effects. In this paper, we address the estimation of Markov and semi-Markov switching linear mixed models in a general framework. We propose a MCEM-like algorithm whose iterations decompose into three steps (sampling of state sequences given random effects, prediction of random effects given the state sequence and maximization). The proposed statistical modeling approach is illustrated by the analysis of successive annual shoots along Corsican pine trunks influenced by climatic covariates.

Key-words: Markov switching model, semi-Markov switching model, individual-wise random effect, MCEM algorithm, plant structure analysis

* CIRAD, UMR Développement et Amélioration des Plantes, Montpellier, France

† INRIA, Équipe-projet Virtual Plants, Montpellier, France

‡ Montpellier III University, Institut de Mathématiques et de Modélisation de Montpellier, Montpellier, France

Combinaisons markoviennes et semi-markoviennes de modèles linéaires mixtes pour identifier les composantes de la croissance d'arbres forestiers

Résumé : La croissance d'un arbre est le résultat de trois composantes: (i) une composante endogène supposée structurée comme une succession de phases stationnaires séparées par des sauts nets asynchrones entre individus, (ii) une composante environnementale pouvant varier dans le temps supposée prendre la forme de fluctuations locales synchrones entre individus, (iii) une composante individuelle qui correspond à l'environnement local de chaque arbre. Afin d'identifier et de caractériser ces trois composantes, nous proposons d'utiliser les combinaisons semi-markoviennes de modèles linéaires mixtes, i.e. des modèles qui combinent des modèles linéaires mixtes de manière semi-markovienne. La semi-chaîne de Markov sous-jacente représente la succession des phases de croissance (composante endogène) tandis que les modèles linéaires mixtes attachés à chaque état de la semi-chaîne de Markov sous-jacente représentent dans la phase de croissance correspondante à la fois l'influence de variables environnementales (composante environnementale) comme des effets fixes et l'hétérogénéité inter-individuelle (composante individuelle) comme des effets alatoires. Dans ce papier, nous traitons de l'estimation des combinaisons markoviennes et semi-markoviennes de modèles linéaires mixtes dans un cadre général. Nous proposons un algorithme de type MCEM dont les itérations se décomposent en trois étapes (simulations des séquences d'états sachant les effets aléatoires, prédiction des effets aléatoires sachant les séquences d'états, maximisation). La modélisation proposée est illustrée par l'analyse des longueurs de pousses annuelles successives le long de tronc de pins Laricio, influencées par des covariables climatiques.

Mots-clés : combinaison markovienne de modèles, combinaison semi-markovienne de modèles, effet aléatoire "individuel", algorithme MCEM, analyse de la structure de la plante

1 Introduction

The analysis of plant structure at macroscopic scales is of major importance in forestry and different fields of agronomy; see Candy (1997), Durand et al. (2005), Guédon et al. (2007) and Ninomiya and Yoshimoto (2008) for illustrations. Tree development can be reconstructed at a given observation date from morphological markers (such as cataphyll[1] or branching scars that help to delimit successive annual shoots) corresponding to past events. Observed growth, as given for instance by the length of successive annual shoots along a tree trunk, is assumed to be mainly the result of three components: an endogenous component, an environmental component and an individual component. The endogenous component is assumed to be structured as a succession of roughly stationary phases that are asynchronous between individuals (Guédon et al., 2007) while the environmental component is assumed to take the form of local fluctuations that are synchronous between individuals. This environmental component is thus assumed to be a "population" component as opposed to the individual component. The environmental factors which modulate the plant development are mainly of climatic origin such as rainfall or temperature. The individual component may cover effects of diverse origins but always includes a genetic effect. Other effects correspond to the local environment of each individual such as pathogen infestation or competitions between trees for light or nutrient resources. These factors are rarely measurable retrospectively for each tree.

Guédon et al. (2007) proposed a set of methods for analyzing the endogenous and the environmental components. In particular, hidden semi-Markov chains with simple observation distributions were applied to forest tree growth data. In this case, the underlying semi-Markov chain represents the succession of growth phases and their lengths while the environmental component is characterized globally. Hidden semi-Markov chains (Guédon, 2003) generalize hidden Markov chains (Ephraim and Merhav, 2002) with the distinctive property of explicitly modeling the sojourn time in each state. Chaubert et al. (2007) applied Markov switching linear mixed models to forest tree growth data. These models combine linear mixed models in a Markovian manner. In this case, the underlying Markov chain represents the succession of growth phases while the linear mixed models attached to each state of the Markov chain represent in the corresponding growth phase both the effect of covariates as fixed effect and inter-individual heterogeneity as random effect.

A Gaussian hidden Markov model can be defined as a pair of stochastic processes $\{S_t, Y_t\}$ where the output process $\{Y_t\}$ is related to the state process $\{S_t\}$, which is a finite-state Markov chain, by the Gaussian distribution $Y_t \mid S_t=s_t \sim \mathcal{N}(\mu_{s_t}, \Gamma^2_{s_t})$. These models, first introduced in speech recognition in the early 1970s, can be viewed as a finite mixture of Gaussian distributions with Markovian dependencies (Ephraim and Merhav, 2002; Cappé et al., 2005). Lindgren (1978) introduced Markov switching linear models which extend the class of Gaussian hidden Markov models by incorporating the influence of covariates as fixed effects in the output process. Markov switching linear models have

---

[1] A reduced or scarcely developed leaf at the start of a plant's life or in the early stages of leaf development.

been applied in such different fields as for instance econometrics or the analysis of gene regulatory networks in biology (Gupta et al., 2007); see Fruhwirth-Schnatter (2006) for an overview of Markov switching models with different examples of application. In the literature, hidden Markov models with random effects in the output process have been used in a limited way. Altman (2007) introduced Markov switching generalized linear mixed model and applied these models for modeling lesion counts in multiple sclerosis patients. Lesion count is assumed to follow a Poisson distribution, with the mean being dependent on the patient's unobserved state. Markov switching generalized linear mixed models were also used in the analysis of symptoms in patients with primary and metastatic brain tumours (Rijmen et al., 2008). Both Altman (2007) and Rijmen et al. (2008) assumed that the individual-wise random effect is independent of the unobservable states.

Here, we introduce semi-Markov switching linear mixed models that generalize both Markov switching linear mixed models and hidden semi-Markov chains. These models can be viewed as a finite mixture of linear mixed models with semi-Markovian dependencies. In a semi-Markov switching linear mixed model applied to forest tree growth data, the underlying semi-Markov chain represents both the succession of growth phases and their lengths, while the linear mixed models attached to each state of the semi-Markov chain represent in the corresponding growth phase the effect of time-varying climatic covariates as fixed effect and inter-individual heterogeneity as random effect. The objective is both to characterize the tree population and to analyse the behavior of each individual within the population. Since studies on plant architecture highlighted the central role played by the endogenous component in plant architecture development (Barthélémy and Caraglio, 2007), a key point is to understand how the effect of climatic factors and inter-individual heterogeneity change with phases.

Since both the states of the underlying (semi-)Markov chain and the random effects are non-observable, (semi-)Markov switching linear mixed models involve two latent structures and remain difficult to estimate. Altman (2007) proposed a deterministic and a stochastic approximation method for estimating Markov switching generalized linear mixed models. The deterministic approximation approach combines numerical integration by Gaussian quadrature and quasi-Newton methods and relies on the fact that the hidden Markov model likelihood can be written as a product of matrices. Since the hidden semi-Markov model likelihood cannot be written as a product of matrices, this deterministic approximation method cannot be transposed to the semi-Markovian case. Moreover, the deterministic approximation approach can only be applied in the case of a few random effects. The stochastic approximation method is a Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990) where the M-step involves quasi-Newton routines. Altman underlined some limitations of the two proposed methods such as the sensitivity to starting values, the slowness to converge and a strong computation burden. Since conditional independence assumptions within a Markov switching linear mixed model can be represented by a directed acyclic graph, Rijmen et al. (2008) proposed to carry out the E-step by a junction tree algorithm (Smyth et al., 1997; Cowell et al., 1999). The M-step involves numerical integration by Gaussian quadrature and Fisher scoring methods. Since conditional independence assumptions within a hidden semi-Markov

model cannot be efficiently represented by a directed acyclic graph, this method cannot be transposed to the semi-Markovian case. Moreover, the approaches proposed by Altman (2007) and Rijmen et al. (2008) cannot be transposed to our context where it is assumed that the random effects are attached to the states. Kim and Smyth (2006) proposed an estimation method for a "left-right" semi-Markov switching linear mixed model with individual-state-wise random effects. Thus, the states are ordered and each state can be visited at most once. The proposed method which is basically an application of the EM algorithm based on a forward-backward algorithm for the E-step relies heavily on the two specific model assumptions (state visited at most once and individual-state-wise random effects). Its complexity is cubic in the sequence length (because of the computation of the marginal observation distributions for each possible state segment location).

We here proposed a MCEM-like algorithm for estimating Markov and semi-Markov switching linear mixed models with either individual-wise or individual-state-wise random effects. Its iterations decompose into three steps: sampling of state sequences given random effects, prediction of random effects given a state sequence and maximization.

This paper is organized as follows. Markov switching linear mixed models are formally defined in Section 2. The maximum likelihood estimation of both Markov and semi-Markov switching linear mixed models with the proposed MCEM-like algorithm is presented in Section 3. The semi-Markov switching linear mixed model is illustrated in Section 4 by the analysis of successive annual shoots along Corsican pine trunks. Section 5 consists of concluding remarks.

2   Model definition

Let $\{S_t\}$ be a Markov chain with finite-state space $\{1, \ldots, J\}$. This J-state Markov chain is defined by the following parameters:

- initial probabilities $\pi_j = P(S_1 = j)$, $j = 1, \ldots, J$ with $\sum_j \pi_j = 1$;

- transition probabilities $p_{ij} = P(S_t = j \mid S_{t-1} = i)$, $i, j = 1, \ldots, J$ with $\sum_j p_{ij} = 1$.

Let Y_{at} be the observation and let S_{at} be the non-observable state for individual a ($a = 1, \ldots, N$), at time t ($t = 1, \ldots, T_a$). Let $\sum_{a=1}^N T_a = T$. $Y_{a1}^{T_a}$ denotes the T_a-dimensional vector of observations on individual a, and $\boldsymbol{Y}$ the T-dimensional vector of all observations. The vectors of non-observable states, $S_{a1}^{T_a}$ and $\boldsymbol{S}$, are defined analogously.

A Markov switching linear mixed model can be viewed as a pair of stochastic processes $\{S_{at}, Y_{at}\}$ where the output process $\{Y_{at}\}$ is related to the state process $\{S_{at}\}$, which is a finite-state Markov chain, by a linear mixed model. We introduce two nested families of Markov switching linear mixed models which differ in the assumptions made concerning inter-individual heterogeneity in the output process:

- Individual-wise random effect:

$$\text{In state } S_{at} = s_{at}, \qquad Y_{at} = X_{at}\beta_{s_{at}} + \tau_{s_{at}}\xi_a + \epsilon_{at}, \qquad (1)$$

$$\xi_a \sim \mathcal{N}(0,1), \qquad \epsilon_{at} \mid S_{at}=s_{at} \sim \mathcal{N}(0,\sigma^2_{s_{at}}).$$

The individual status (compared to the average individual) within the population is common to all the states.

- Individual-state-wise random effect:

$$\text{In state } S_{at} = s_{at}, \qquad Y_{at} = X_{at}\beta_{s_{at}} + \tau_{s_{at}}\xi_{as_{at}} + \epsilon_{at}, \qquad (2)$$

$$\xi_{as_{at}} \sim \mathcal{N}(0,1), \qquad \epsilon_{at} \mid S_{at}=s_{at} \sim \mathcal{N}(0,\sigma^2_{s_{at}}).$$

The individual status is different in each state.

In these definitions, X_{at} is the Q-dimensional row vector of covariates for individual a at time t, ξ_a is the individual a random effect. Given the state $S_{at} = s_{at}$, $\beta_{s_{at}}$ is the Q-dimensional fixed effect parameter vector, $\xi_{as_{at}}$ is the individual a random effect, $\tau_{s_{at}}$ is the standard deviation for the random effect and $\sigma^2_{s_{at}}$ is the residual variance. The individuals are assumed to be independent. For convenience, random effects are supposed to follow the standard Gaussian distribution. In the individual-state-wise random effect model, the random effects for an individual a are assumed to be independent between the non-observable states $(\text{cov}(\xi_{aj}, \xi_{aj'}) = 0; j \neq j')$. Including random effects in the output process cancels the assumption that the successive observations for an individual are conditionally independent given the non-observable states. The successive observations for an individual are here assumed to be conditionally independent given the non-observable states and random effects. In state j, the introduction of random effects makes it possible to decompose the total variance Γ_j^2 into two parts: variance due to inter-individual heterogeneity τ_j^2 and residual variance σ_j^2 as $\Gamma_j^2 = \tau_j^2 + \sigma_j^2$.

3   Maximum likelihood estimation with a Monte Carlo EM-like algorithm

The Markov switching linear mixed model parameters can be divided into two categories: the parameters $\pi = (\pi_j; j = 1, \ldots, J)$ and $P = (p_{ij}; i, j = 1, \ldots, J)$ of the underlying Markov chain and the parameters $\beta = (\beta_j; j = 1, \ldots, J)$, $\tau = (\tau_j; j = 1, \ldots, J)$ and $\sigma^2 = (\sigma_j^2; j = 1, \ldots, J)$ of the J linear mixed models. In the following, we denote by $\theta = (\pi, P, \beta, \tau, \sigma^2)$ the set of parameters to be estimated. The maximum likelihood estimation is presented in the case of the individual-state-wise random effect model. The transposition to the individual-wise random effect model is straightforward.

Let $\xi_{a1}^J = (\xi_{aj}; j = 1, \ldots, J)$ be the J-dimensional random effect vector for individual a. The likelihood function of the observed data is given by:

$$
\begin{aligned}
L(\boldsymbol{y};\theta) &= \prod_{a=1}^{N}\int_{\xi_{a1}^{J}}\left\{\sum_{s_{a1}^{T_a}}f(s_{a1}^{T_a},\xi_{a1}^{J},y_{a1}^{T_a};\theta)\right\}d\xi_{a1}^{J} \\
&= \prod_{a=1}^{N}\int_{\xi_{a1}^{J}}\left\{\sum_{s_{a1}^{T_a}}f(s_{a1}^{T_a};\theta)f(\xi_{a1}^{J};\theta)f(y_{a1}^{T_a}\mid s_{a1}^{T_a},\xi_{a1}^{J};\theta)\right\}d\xi_{a1}^{J}, \quad (3)
\end{aligned}
$$

where $\sum_{s_{a1}^{T_a}}$ means "sum on all possible state sequences of length T_a for individual a". Since the likelihood function (Equation 3) cannot be evaluated in closed form, the direct maximization of this likelihood function is difficult.

Since both the states of the underlying Markov chain and the random effects are non observable, the EM algorithm (McLachlan and Krishnan, 2008) appears at first sight as a natural candidate to estimate Markov switching linear mixed models. Let us consider the complete-data log-likelihood where both the outputs $\boldsymbol{y}$, the random effects $\boldsymbol{\xi}$ and the states $\boldsymbol{s}$ of the underlying Markov chain are observed

$$
\begin{aligned}
\log f(\boldsymbol{s},\boldsymbol{\xi},\boldsymbol{y};\theta) &= \sum_{a=1}^{N}\log f(s_{a1}^{T_a},\xi_{a1}^{J},y_{a1}^{T_a};\theta) \\
&= \sum_{a=1}^{N}\left\{\log f(s_{a1}^{T_a};\theta)+\log f(\xi_{a1}^{J};\theta)+\log f(y_{a1}^{T_a}\mid s_{a1}^{T_a},\xi_{a1}^{J};\theta)\right\} \\
&= \sum_{a=1}^{N}\sum_{j=1}^{J}I\Big(s_{a1}=j\Big)\log\pi_j \\
&\quad +\sum_{a=1}^{N}\sum_{t=2}^{T_a}\sum_{i,j=1}^{J}I\Big(s_{at}=j,s_{a,t-1}=i\Big)\log p_{ij} \\
&\quad +\sum_{a=1}^{N}\sum_{j=1}^{J}\log\phi(\xi_{aj};0,1) \\
&\quad +\sum_{a=1}^{N}\sum_{t=1}^{T_a}\sum_{j=1}^{J}I\Big(s_{at}=j\Big)\log\phi(y_{at};X_{at}\beta_j+\tau_j\xi_{aj},\sigma_j^2), \quad (4)
\end{aligned}
$$

where $\phi(y;\mu,\sigma^2)$ is the Gaussian density with mean μ and variance σ^2 and $I()$ is the indicator function.

The E-step of the EM algorithm requires calculating the conditional expectation of $\log f(\boldsymbol{s},\boldsymbol{\xi},\boldsymbol{y};\theta)$ given the observed data $\boldsymbol{y}$ and the current value of θ. But, the EM algorithm for hidden Markov models (Ephraim and Merhav, 2002) cannot be transposed because the successive observations for an individual are not conditionally independent given the non-observable states; see Section 2. The EM algorithm for finite mixture of linear mixed models (Celeux et al., 2005) cannot be adapted because the distributions of $\xi_{aj}\mid Y_{a1}^{T_a}=y_{a1}^{T_a}$ and $\xi_{aj}\mid S_{at}=j,Y_{a1}^{T_a}=y_{a1}^{T_a}$ cannot be analytically derived for each individual a at time t. The Monte Carlo EM (MCEM) algorithm (Wei and Tanner, 1990), where the quantities computed in the deterministic E-step are approximated using Monte Carlo methods, is the usual alternative solution.

For the presentation of the estimation algorithm, we adopted the framework of restoration-maximization (RM) algorithms proposed by Qian and Titterington (1991) (see also Archer and Titterington (2002)). The Monte Carlo EM algorithm proposed by Altman (2007) can be interpreted as a RM algorithm with two restoration steps for the two latent structures, a unconditional stochastic one for the random effects and a conditional deterministic one for the state sequences (the unconditional/conditional qualifier refers to the other latent structure). We did not adopt a similar approach since, in our definition of Markov switching linear mixed models (see Section 2), the random effects are attached to the states. Hence, following Shi and Lee (2000), we rather chose to perform two conditional restoration steps, one for the random effects given the state sequences (and the observed data) and one for the state sequences given the random effects (and the observed data). The iteration of the proposed RM algorithm decomposes in:

- Conditional R-step for state sequences:
  for each individual a, sample $s_{a1}^{T_a}$ from $P(S_{a1}^{T_a} = s_{a1}^{T_a} \mid \xi_{a1}^J, Y_{a1}^{T_a} = y_{a1}^{T_a}; \theta)$ by a direct transposition of the forward-backward algorithm proposed by Chib (1996),

- Conditional R-step for random effects:
  for each individual a, compute the best posterior prediction ξ_{a1}^J from $P(\xi_{a1}^J \mid S_{a1}^{T_a} = s_{a1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}; \theta)$,

- Maximisation-step.

As discussed by Neal and Hinton (1998) in the Gaussian mixture case, since Markov chain parameters and linear mixed model parameters form disjoint sets and influence the complete-data log-likelihood separately (Equation 4), Markov chain parameters can be updated when the distribution of $\boldsymbol{S}$ is re-calculated and linear mixed model parameters can be updated when the $\boldsymbol{\xi}$ are predicted. It makes sense to immediately re-estimate the parameters before performing the conditional R-step for the other latent structure. This approach, called incremental EM algorithm, permits to speed up the convergence. It can be noted that the order of the steps is not important and does not influence the parameters estimation and the convergence of the algorithm.

3.1 Forward-backward algorithm for sampling state sequences given the random effects

For each individual a, the state sequences are sampled from the conditional distribution $P(S_{a1}^{T_a} = s_{a1}^{T_a} \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J)$.

For a Markov switching linear mixed model, since

$$
\begin{aligned}
&P\big(S_{a1}^{T_a} = s_{a1}^{T_a} \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) \\
=\ &\left\{ \prod_{t=1}^{T_a-1} P\big(S_{at} = s_{at} \mid S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) \right\} \\
\times\ &P\big(S_{aT_a} = s_{aT_a} \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big),
\end{aligned}
$$

the following conditional distributions should be used for sampling state sequences:

- final state (initialization)

$$P\big(S_{aT_a} = s_{aT_a} \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big),$$

- previous state

$$P\big(S_{at} = s_{at} \mid S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big).$$

The forward-backward algorithm for sampling state sequences given the random effects can be decomposed into two passes, a forward recursion which is similar to the forward recursion of the forward-backward algorithm for hidden Markov chains, and a backward pass for sampling state sequences (Chib, 1996); see the Appendix for details.

The forward recursion can be used to compute the observed-data log-likelihood given the random effects:

$$
\begin{aligned}
\log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}; \theta) &= \sum_{a=1}^{N} \Bigg(\log P(Y_{a1} = y_{a1} \mid \xi_{a1}^J; \theta) \\
&\qquad + \sum_{t=2}^{T_a} \log P(Y_{at} = y_{at} \mid Y_{a1}^{t-1} = y_{a1}^{t-1}, \xi_{a1}^J; \theta) \Bigg) \\
&= \sum_{a=1}^{N} \sum_{t=1}^{T_a} \log N_{at}. \qquad (5)
\end{aligned}
$$

where N_{at} is the normalizing factor for individual a at time t; see Appendix.

3.2   Random effect prediction given the state sequence

The predicted vector for the random effects ξ_{a1}^J attached to individual a is:

$$
\begin{aligned}
\xi_{a1}^J(m) &= \mathrm{E}\Big[\xi_{a1}^J \mid S_{a1}^{T_a} = s_{a1}^{T_a}(m),\ Y_{a1}^{T_a} = y_{a1}^{T_a}\Big] \qquad (6) \\
&= \Omega U_a^{(m)\prime}\Big(U_a^{(m)}\Omega^2 U_a^{(m)\prime} + \mathrm{Diag}\{U_a^{(m)}\sigma^2\}\Big)^{-1}\Big(y_{a1}^{T_a} - \sum_{j=1}^{J} I_{aj}(m)X_a\beta_j\Big),
\end{aligned}
$$

where

- $s_{a1}^{T_a}(m)$ is the mth state sequence sampled for individual a,

- $\Omega = \mathrm{Diag}\{\tau_j; j = 1, \ldots, J\}$ is the $J \times J$ random standard deviation matrix,

- $U_a^{(m)}$ is the $T_a \times J$ design matrix associated with state sequence $s_{a1}^{T_a}(m)$, composed of 1 and 0 with $\sum_j U_a^{(m)}(t,j) = 1$ and $\sum_t \sum_j U_a^{(m)}(t,j) = T_a$,

- $u_{at}^{(m)} = \Big(I(s_{at}(m) = 1) \cdots I(s_{at}(m) = J)\Big)$ is the tth row of the design matrix $U_a^{(m)}$,

- $\sigma^2 = (\sigma_1^2 \cdots \sigma_J^2)'$ is the J-dimensional residual variance vector,

- Diag$\{U_a^{(m)}\sigma^2\}$ is the $T_a \times T_a$ diagonal matrix with $\{u_{at}^{(m)}\sigma^2; t = 1,\ldots,T_a\}$ on its diagonal,

- $I_{aj}(m) = \text{Diag}\Big\{I(s_{at}(m) = j), \ t = 1,\cdots,T_a\Big\}$ is a $T_a \times T_a$ diagonal matrix,

- X_a is the $T_a \times Q$ matrix of covariates.

3.3 M-step

In the proposed MCEM-like algorithm, the conditional expectation of the complete-data log-likelihood given the observed data is approximated at iteration k by:

$$
\begin{aligned}
\text{E}\Big[\log f(\boldsymbol{s}, \boldsymbol{\xi}, \boldsymbol{y}; \theta) \mid \boldsymbol{Y} = \boldsymbol{y}; \theta^{(k)}\Big] &= \sum_{a=1}^{N} \text{E}\Big[\log f(s_{a1}^{T_a}, \xi_{a1}^{J}, y_{a1}^{T_a}; \theta) \mid Y_{a1}^{T_a} = y_{a1}^{T_a}; \theta^{(k)}\Big] \\
&\approx \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\log f\Big(s_{a1}^{T_a(k)}(m), \xi_{a1}^{J(k)}(m), y_{a1}^{T_a}; \theta^{(k)}\Big) \\
&\approx \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{j=1}^{J} I\Big(s_{a1}^{(k)}(m) = j\Big)\log \pi_j^{(k)} \\
&+ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{t=2}^{T_a}\sum_{i,j=1}^{J} I\Big(s_{at}^{(k)}(m) = j, \ s_{a,t-1}^{(k)}(m) = i\Big)\log p_{ij}^{(k)} \\
&+ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{j=1}^{J}\log \phi(\xi_{aj}^{(k)}(m); 0, 1) \\
&+ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{t=1}^{T_a}\sum_{j=1}^{J} I\Big(s_{at}^{(k)}(m) = j\Big)\log \phi(y_{at}; X_{at}\beta_j^{(k)} + \tau_j^{(k)}\xi_{aj}^{(k)}(m), \sigma_j^{2(k)}).
\end{aligned}
\tag{7}
$$

where M_k is the number of state sequences sampled at iteration k for each individual a.

At iteration k, the new values of the parameters of the Markov switching linear mixed model are obtained by maximizing the different terms of Equation 7, each term depending on a given subset of θ.

For the parameters of the underlying Markov chain, we obtain:

- initial probabilities

$$
\pi_j^{(k+1)} = \frac{\sum_a\sum_m I\Big(s_{a1}^{(k)}(m) = j\Big)}{NM_k},
$$

- transition probabilities

$$
p_{ij}^{(k+1)} = \frac{\sum_a\sum_m\sum_{t=2}^{T_a} I\Big(s_{at}^{(k)}(m) = j, \ s_{a,t-1}^{(k)}(m) = i\Big)}{\sum_a\sum_m\sum_{t=2}^{T_a} I\Big(s_{a,t-1}^{(k)}(m) = i\Big)}.
$$

For the parameters of the J linear mixed models, we obtain:

- fixed effect parameters

$$\beta_j^{(k+1)} = \left(\sum_a \sum_m X_a' I_{aj}^{(k)}(m) X_a \right)^{-1} \left(\sum_a \sum_m X_a' I_{aj}^{(k)}(m) \left(y_{a1}^{T_a} - \tau_j^{(k)} \xi_{aj}^{(k)}(m) \right) \right), \quad (8)$$

- residual variances

$$\sigma_j^{2(k+1)} = \frac{\sum_a \sum_m \left(y_{a1}^{T_a} - X_a \beta_j^{(k)} - \tau_j^{(k)} \xi_{aj}^{(k)}(m) \right)' I_{aj}^{(k)}(m) \left(y_{a1}^{T_a} - X_a \beta_j^{(k)} - \tau_j^{(k)} \xi_{aj}^{(k)}(m) \right)}{\sum_a \sum_m \operatorname{tr}\{I_{aj}^{(k)}(m)\}}, \quad (9)$$

- random effect standard deviations

$$\tau_j^{(k+1)} = \frac{\sum_a \sum_m \sum_t I\left(s_{at}^{(k)}(m) = j\right) \xi_{aj}^{(k)}(m) \left(y_{at} - X_{at}\beta_j^{(k)}\right)^2}{\sum_a \sum_m \sum_t I\left(s_{at}^{(k)}(m) = j\right) \xi_{aj}^{2(k)}(m)}. \quad (10)$$

The reestimation of linear mixed model parameters is similar to the reestimation of linear model parameters by ordinary least squares. The difference is that the new values of the linear mixed model parameters are weighted by the number of occurrence of states within sampled state sequences.

3.4   Initialisation of the algorithm

Various simulations have been conducted using different starting values. The proposed MCEM-like algorithm is indeed sensitive to starting values. The more distant from true values the starting values are, the worse the parameter estimates are. We recommend to choose as starting values the parameters estimated by the EM algorithm for a simple Markov switching linear model or semi-Markov switching linear model.

3.5   Sample size

The conditional R-steps relies on the restoration of several pairs $(s_{a1}^{T_a}, \xi_{a1}^J)$ for each individual a. As discussed by Wei and Tanner (1990) and Cappé et al. (2005), it is inefficient to start with a large number of sampled state sequences M_k. They recommended to increase M_k as the current approximation moves closer to the true maximizer. In order not to increase exponentially the number of sampling pairs over the iterations, we propose to introduce a further step in the RM algorithm proposed at iteration k:

- Conditional R-step for state sequences:
  sample M_k state sequences from $P(S_{a1}^{T_a} = s_{a1}^{T_a} \mid \xi_{a1}^J, Y_{a1}^{T_a} = y_{a1}^{T_a}; \theta)$ for each individual a,

- Conditional R-step for random effects:
  predict a random effect from $P(\xi_{a1}^J \mid S_{a1}^{T_a} = s_{a1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}; \theta)$ for each individual a and each sampled state sequences,

- Maximisation-step,

- Choice of sample size-step:
  increase the number of sampled pairs M_k and sample with replacement M_{k+1} random effects among the M_k predicted random effects.

Caffo et al. (2005) proposed an adaptative strategy in order to choose the sample size and to recover the ascent property (increase of the observed data likelihood) of the EM algorithm.

3.6   Convergence of the algorithm

Due to the sample variability introduced by Monte Carlo step, the ascent property of the EM algorithm is lost (McLachlan and Krishnan, 2008). To determine the convergence of the MCEM algorithm, Wei and Tanner (1990) recommended to monitor the plots of $\theta^{(k)}$ against the iteration index k and to terminate when the plots exhibit random fluctuations around a roughly stationary value θ^*. Another way is to monitor the convergence by the difference between consecutive observed likelihoods. However, if the observed data likelihood cannot be evaluated analytically, Meng and Schilling (1996) suggested to monitor the convergence by the bridge sampling method that can be used to obtain an estimate of the observed data likelihood ratio between two consecutive iterations. Nevertheless, in the Markov switching linear mixed model case, the observed data likelihood cannot be computed at each iteration.

Under assumption of convergence of random effect predictions, we chose to monitor the convergence of the proposed MCEM algorithm by the difference between successive iterations

$$\log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}^{(k+1)}; \theta^{(k+1)}) - \log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}^{(k)}; \theta^{(k)}) \tag{11}$$

where the quantity $\log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}^{(k)}; \theta^{(k)})$ is directly obtained as a byproduct of the forward recursion (Equation 5).

3.7   MCEM-like algorithm for individual-wise random effect models

The transposition to individual-wise random effect models is straightforward. Since the individual-wise random effects are incorporated in the output process, the main difference concerns the conditional R-step of random effect prediction given a state sequence (Equation 6). In the M-step (Equations 8, 9 and 10) and in the forward and backward passes (Section 3.1), the random effects ξ_{a1}^J are replaced by ξ_a. Using the notations introduced in Section 3.2, the predicted random effect ξ_a for each individual a is given by:

$$
\begin{aligned}
\xi_a(m) &= \mathrm{E}\Big[\xi_a \mid S_{a1}^{T_a} = s_{a1}^{T_a}(m), Y_{a1}^{T_a} = y_{a1}^{T_a}\Big] \\
&= \tau' U_a^{(m)\prime}\Big(U_a^{(m)}\tau\tau' U_a^{(m)\prime} + \mathrm{Diag}\{U_a^{(m)}\sigma^2\}\Big)^{-1}\Big(y_{a1}^{T_a} - \sum_{j=1}^{J} I_{aj}(m) X_a \beta_j\Big),
\end{aligned}
$$

where $\tau = (\tau_1 \cdots \tau_J)'$ is the J-dimensional random effect standard deviation vector.

3.8 Extension to semi-Markov switching linear mixed models

Semi-Markov chains generalize Markov chains with the distinctive property of explicitly modeling the sojourn time in each state. Let $\{S_t\}$ be a finite-state semi-Markov chain defined by the following parameters:

- initial probabilities $\pi_j = P(S_1 = j)$, with $\sum_j \pi_j = 1$,

- transition probabilities

    - nonabsorbing state i:
      for each $j \neq i$, $\tilde{p}_{ij} = P(S_t = j \mid S_t \neq i, S_{t-1} = i)$, with $\sum_{j \neq i} \tilde{p}_{ij} = 1$ and $\tilde{p}_{ii} = 0$,

    - absorbing state i:
      $p_{ii} = P(S_t = i \mid S_{t-1} = i) = 1$ and for each $j \neq i$, $p_{ij} = 0$.

An occupancy distribution is attached to each nonabsorbing states; $u = 1, 2, \ldots$:

$$d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \ldots, u-2 \mid S_{t+1} = j, S_t \neq j). \quad (12)$$

As for a Markov switching linear mixed model, the output process $\{Y_{at}\}$ of the semi-Markov switching linear mixed model for individual a is related to the underlying semi-Markov chain $\{S_{at}\}$ by the linear mixed model (1) in the case of individual-wise random effects or by the linear mixed model (2) in the case of individual-state-wise random effects. Since covariates and random effects are incorporated in the output process, the successive observations for an individual are assumed to be conditionally independent given the non-observable states and the random effects. The proposed MCEM-like algorithm can therefore be directly transposed to semi-Markov switching linear mixed models. Given the random effects, the state sequences are sampled using the forward-backward algorithm adapted to hidden semi-Markov chains (see Guédon (2007) and references therein). Given a state sequence, the random effects are predicted as previously described. The underlying semi-Markov chain parameters (initial probabilities, transition probabilities and state occupancy distributions) and the linear mixed model parameters (fixed effect parameters, random variance and residual variance) are obtained by maximizing the Monte Carlo approximation of the conditional expectation of the complete-data log-likelihood. The reestimation of the initial probabilities, the transition probabilities and the state occupancy distributions (M-step of the MCEM algorithm) is similar to the reestimation in the hidden semi-Markov chain case derived by Guédon (2003), the smoothed probabilities being simply replaced by counting.

4 Application to Corsican pine growth

The use of semi-Markov switching linear mixed models is illustrated by the analysis of forest tree growth. The data set comprised four sub-samples of Corsican pines planted in a forest stand in the "Centre" region (France) : 31 6-year-old trees, 29 12-year-old trees (first year not measured), 30 18-year-old trees (first year not measured) and 13 23-year-old trees (two first years not measured). Trees in the first sub-sample (6-year-old) remained in the nursery for two years

before transplantation while trees in the three other sub-samples remained in the nursery for three years before transplantation. Trees trunks were described by annual shoot from the top to the base where the length (in cm) was recorded for each annual shoot (Figure 1). The annual shoot is defined as the segment of stem established within a year. The observed growth is mainly the result of the modulation of the endogenous growth component by climatic and individual factors. The endogenous growth component is assumed to be structured as a succession of roughly stationary phases separated by marked change points that are asynchronous between trees (Guédon et al., 2007). Trees were chosen in order to cover the entire range of behaviors and were not subject to any silvi-cultural interventions. The length of successive annual shoots along tree trunks was previously analyzed using a hidden semi-Markov chain (Guédon et al., 2007) and a Markov switching linear mixed model (Chaubert et al., 2007). In the first case, the effect of climatic factors and the inter-individual heterogeneity were not explicitly modeled while in the second case, the length of the successive growth phases was not explicitly modeled.
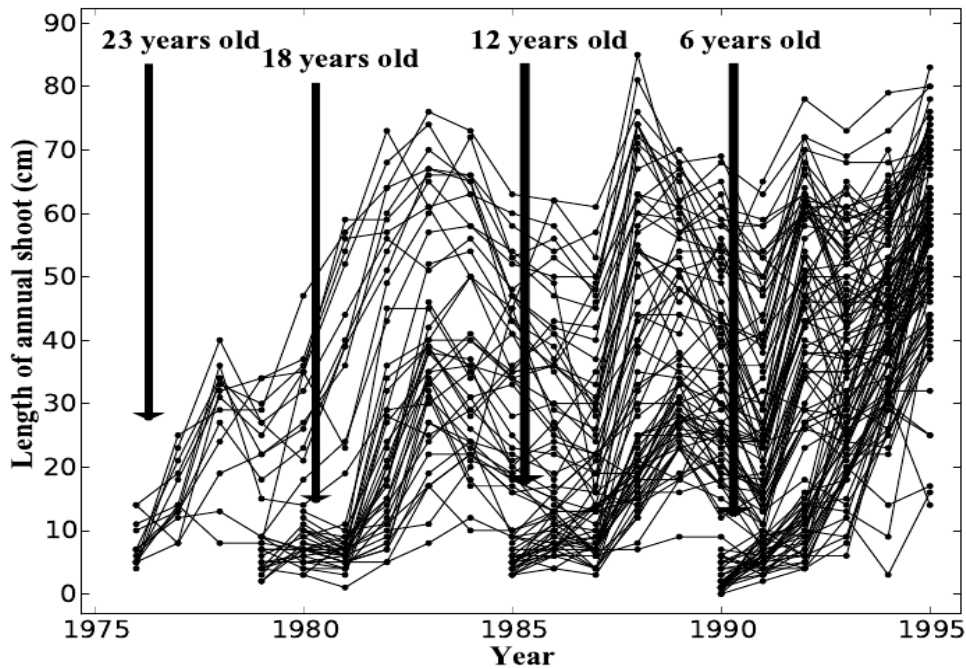

Figure 1: Four sub-samples of Corsican pines: Length of successive annual shoots along tree trunks.

A "left-right" three-state semi-Markov switching linear mixed model com-posed of two successive transient states followed by a final absorbing state was estimated. In temperate regions, rainfall can have a one-year-delayed effect (on the number of leaves) or an immediate effect (on shoot elongation) depending

on whether it occurs during organogenesis[2] or elongation. We chose to use an intercept and the centered cumulated rainfall (in mm) during a period covering one organogenesis period and one elongation period as fixed effects for each linear mixed model. The linear mixed model attached to state j is:

$$y_{at} \mid {}_{S_{at}=j} = \beta_{j1} + \beta_{j2} X_t + \tau_j \xi_{aj} + \epsilon_{at}, \quad \xi_{aj} \sim \mathcal{N}(0,1), \quad \epsilon_{at} \mid {}_{S_{at}=j} \sim \mathcal{N}(0, \sigma_j^2),$$

where y_{at} is the length of the annual shoot for individual a at time t, β_{j1} is the intercept, X_t is the centered cumulated rainfall at time t ($\mathrm{E}(X_t) = 0$) and β_{j2} is the cumulated rainfall parameter. Because of the centering of climatic covariate, the intercept β_{j1} is directly interpretable as the average length of successive annual shoots within state j.
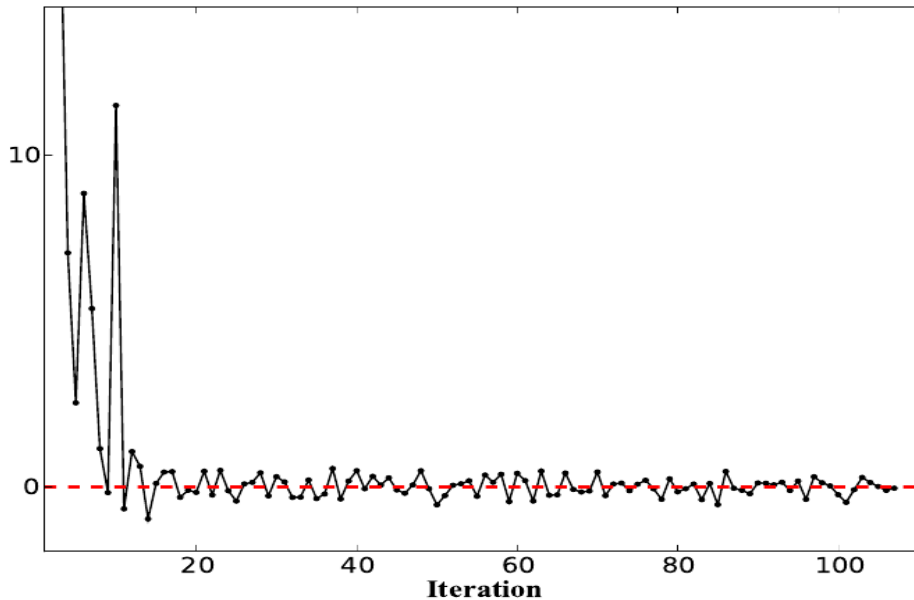
Figure 2: Values of the difference of observed-data log-likelihood given random effects between successive iterations (Equation 11) for the estimated semi-Markov switching linear mixed model with individual-state-wise random effects.

The estimation algorithm was initialized with the parameters π, P, d, β and σ^2 estimated without taking into account the random effects (hence, $\boldsymbol{\xi} = 0$). Once the random effects had converged, the convergence of the algorithm was monitored by the difference between two consecutive iterations of the observed data log-likelihood given the random effects (Equation 11). The plot of the values against the iteration rank for the semi-Markov switching linear mixed model showed that the estimation algorithm converged rapidly by, say, about 70 iterations (Figure 2) with $M_k = k$ state sequences sampled for each tree at the kth iteration. States 1 and 2 are the only possible states (with $\widehat{\pi}_1 = 0.95$ and

[2]The origin and development of the organs and systems of organs in plants.

$\widehat{\pi}_2 = 0.05$) of the estimated underlying semi-Markov chain; see Figure 3. The estimated transition probability matrix is degenerated i.e. for each transient state i, $p_{i,i+1} = 1$ and $p_{ij} = 0$ for $j \neq i+1$ (and for the final absorbing state $p_{ii} = 1$ and $p_{ij} = 0$ for $j \neq i$). It should be noted that the succession of states is deterministic for a degenerated "left-right" semi-Markov switching linear mixed model. This deterministic succession of states supports the assumption of a succession of growth phases.
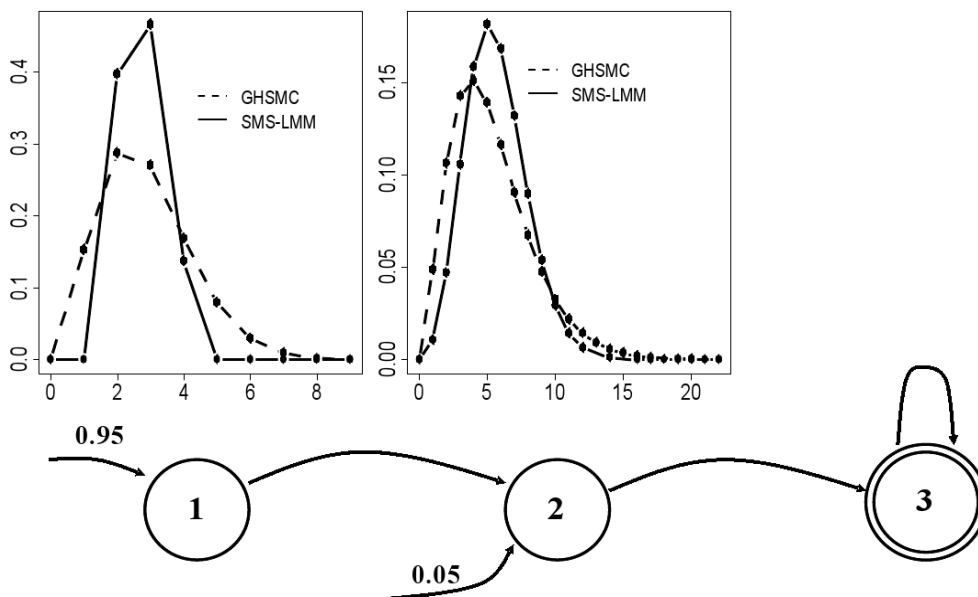
Figure 3: Estimated underlying semi-Markov chain. Each state is represented by a vertex which is numbered. Vertices representing transient states are edged by a single line while the vertex representing the final absorbing state is edged by a double line. Possible transitions between states are represented by arcs (attached probabilities always equal to 1 are not shown). Arcs entering in states indicate initial states. The attached initial probabilities are noted nearby. The occupancy distributions of the nonabsorbing states are shown above the corresponding vertices. The dotted lines correspond to occupancy distributions estimated by a simple Gaussian hidden semi-Markov chain (GHSMC) and the point lines correspond to occupancy distributions estimated by a semi-Markov switching linear mixed model (SMS-LMM).

The marginal observation distribution of the linear mixed model attached to state j is the Gaussian distribution $\mathcal{N}(\mu_j, \Gamma_j^2)$ with $\mu_j = \beta_{j1} + \beta_{j2}\mathrm{E}_j(X)$ and $\Gamma_j^2 = \tau_j^2 + \sigma_j^2$, where $\mathrm{E}_j(X)$ is the mean centered cumulated rainfall X in state j. The marginal observation distribution represents the length of the annual shoots of the average tree in state j. The marginal observation distributions for the different states are well separated (little overlapping between marginal observation distributions corresponding to two successive states); compare the mean difference $\mu_{j+1} - \mu_j$ between consecutive states with the standard deviations Γ_j and Γ_{j+1} in Table 1.

		State j		
		1	2	3
Occupancy distributions mean, sd	GHSMC	$2.88, 1.37$	$5.31, 2.93$	
	SMS-LMM	$2.73, 0.68$	$5.56, 2.20$	
Regression parameters (SMS-LMM)	Intercept β_{j1}	7.09	25.79	50.25
	Cumulated rainfall parameter β_{j2}	0.0027	0.0165	0.0309
	Average cumulated rainfall effect $\beta_{j2} \times sd_j(X)$	0.30	2.16	4.52
Variability decomposition (SMS-LMM)	Random variance τ_j^2	5.79	49.89	69.39
	Residual variance σ_j^2	4.74	39.95	76.86
	Total variance Γ_j^2	10.53	89.84	146.25
	Proportion of inter-individual heterogeneity	54.99%	55.53%	47.45%
Marginal observation distribution μ_j, Γ_j	GHSMC	$6.97, 3.26$	$26.30, 9.12$	$54.35, 11.39$
	SMS-LMM	$6.99, 3.24$	$25.88, 9.48$	$50.32, 12.09$

Table 1: Comparison of the estimated Gaussian hidden semi-Markov chain (GHSMC) parameters with estimated semi-Markov switching linear mixed model (SMS-LMM) parameters (state occupancy distributions and marginal observation distributions). For each observation linear mixed model, the regression parameters, the cumulated rainfall effect and the variability decomposition are given.

The average cumulated rainfall effect (i.e. the average amplitude of the climatic fluctuations) was computed as $\beta_{j2} \times sd_j(X)$ for each state j where $sd_j(X)$ is the standard deviation of the centered cumulated rainfall in state j. The effect of cumulated rainfall was weak in the first state (of slowest growth) while it was strong in the last two states (a little less in the second state than in the third state); see Table 1. The proportion of inter-individual heterogeneity, defined by the ratio between the random variance τ_j^2 and the total variance Γ_j^2 in state j, was greater in early plant life (first two states with more than 54%) and decreased slightly in the last state (near 47%).

At the last iteration of the algorithm, the median predicted random effects were computed for each individual based on the predicted random effects for each individual. The most probable state sequence given the median predicted individual-state-wise random effects was computed for each observed sequence using a Viterbi-like algorithm (Guédon, 2003). The restored state sequence can be viewed as the optimal segmentation of the corresponding observed sequence into sub-sequences, each corresponding to a given state. The fixed part of the three observation linear mixed models (i.e. $\beta_{j1} + \beta_{j2}X_t$ for each state j) for 18-year-old and 23-year-old trees is shown in Figure 4. This confirms that the states are well-separated with little overlapping and correspond to a growth increase.
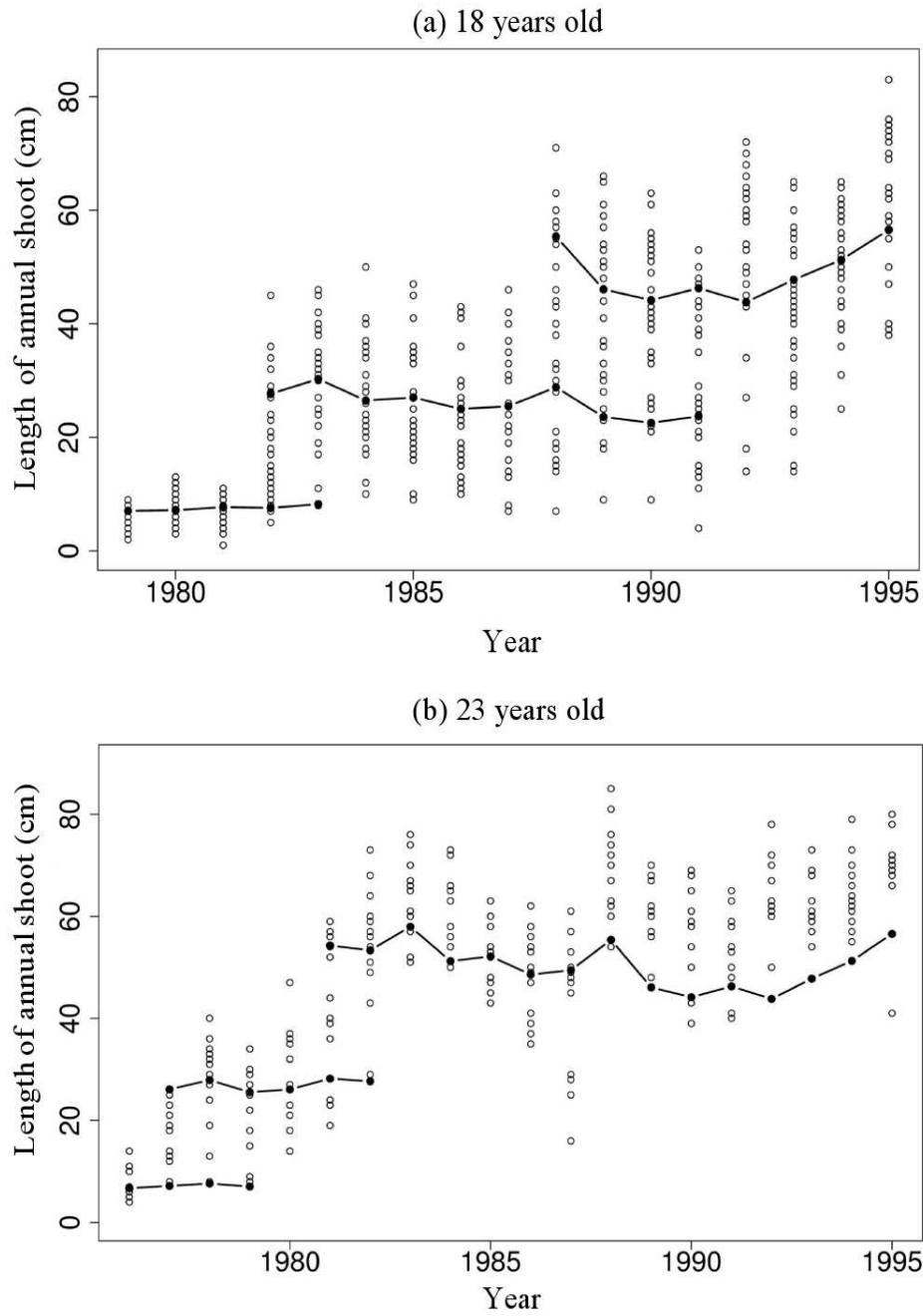
Figure 4: Fixed part of the three observation linear mixed models (i.e. $\beta_{j1} + \beta_{j2}X_t$ for each state j) represented by point lines and observed annual shoot lengths (points): (a) 18-year-old trees, (b) 23-year-old trees.

The characteristics (median and dispersion) of the first year in each state were extracted for the four sub-samples of Corsican pines on the basis of the most probable state sequences of the observed sequences computed using the estimated Gaussian hidden semi-Markov chain and semi-Markov switching linear mixed model. The median first year in the second state for the four sub-samples was similar for the two models; see Table 2. The median first year in the third state for the 6-year-old Corsican pines was similar for the two models. A shift of one year was noted for the median first year in the third state between the two models for the 12-year-old, 18-year-old and 23-year-old Corsican pines. The dispersion of the first year in the second and the third state was greatly reduced in the semi-Markov switching linear mixed model case compared to the Gaussian hidden semi-Markov chain case (Table 2). When the effect of climatic covariates and inter-individual heterogeneity were taken into account, this rendered the states more synchronous between individuals; see also the estimated occupancy distributions (Equation 12), and in particular their standard deviations, for the Gaussian hidden semi-Markov chain and for the semi-Markov switching linear mixed model in Table 1 and Figure 3.

		6 years old	12 years old	18 years old	23 years old
State 2	GHSMC	1993 (0.56)	1988 (0.42)	1982 (1.93)	1978 (1.04)
	SMS-LMM	1993 (0.56)	1988 (0.50)	1982 (0.56)	1978 (1.17)
State 3	GHSMC	1995 (0.35)	1993 (1.00)	1988 (2.56)	1981 (0.78)
	SMS-LMM	1995 (0.38)	1992 (0.87)	1989 (1.27)	1982 (0.77)

Table 2: Median first year in state 2 and state 3 for each sub-sample deduced from the estimated Gaussian hidden semi-Markov chain (GHSMC) and the semi-Markov switching linear mixed model (SMS-LMM). The corresponding standard deviations are indicated in brackets.

The correlation coefficient between the median predicted random effect in state 1 and the median predicted random effect in state 2 was 0.26 while the correlation coefficient between the median predicted random effect in state 2 and the median predicted random effect in state 3 was 0.62. Hence, the behavior of an individual is more strongly related between the last two states than between the first two states. The more general assumption of the individual-state-wise random effect model (a random effect attached to each state) compared to individual-wise random effect model (a random effect common to all states) is more representative of Corsican pine behavior; see Chaubert et al. (2007) for the same conclusion in the Markov switching linear mixed model case.

Complementary biological results about Corsican pine growth and other forest tree species (sessile oaks and Scots pines) can be found in Chaubert-Pereira et al. (2008).

5   Concluding remarks

Semi-Markov switching linear mixed models enable to identify and to characterize the different growth components (endogenous, environmental and individual components) of forest trees. The introduction of climatic covariates and individual-state-wise random effects renders the endogenous growth component more synchronous between individuals than with a simple Gaussian hidden semi-Markov chain. The growth phases are thus not only defined by the averaged length of annual shoots but also by the amplitude of fluctuations synchronous between individuals. Moreover, the behavior of each tree within the population can be investigated on the basis of the predicted individual-state-wise random effects.

The linear mixed model associated with the third state seems to underestimate the observed mean length of successive annual shoots in this phase for the 23-year-old Corsican pines (Figure 4). This behaviour highlights a subsample or group effect. A possible extension of the observation linear mixed model would be to incorporate a group-wise random effect in addition to the individual-state-wise random effect and the fixed effect. It can be noted that incorporating a group-wise random effect can be useful in different situations. The difference between groups can have various causes; for instance, genetic factors, age, plot density or soil properties.

In the proposed MCEM-like algorithm, the conditional restoration step for state sequences given random effects relies on simulations while the conditional restoration step for random effects given state sequences is deterministic. In this latter case, an alternative solution would be to sample random effects applying a Metropolis-Hastings algorithm; see McCulloch (1997) in the generalized linear mixed model case and Lavergne et al. (2007) in the case of mixture of generalized linear mixed models.

The estimation algorithms proposed in this paper can directly be transposed to other families of hidden Markov models such as for instance hidden Markov tree models; see Durand et al. (2005) and references therein. Another interesting direction for further research would be to develop the statistical methodology for semi-Markov switching generalized linear mixed models to take into account non-normally distributed response variables (for instance, number of growth units, apex death/life, non flowering/flowering character in the plant architecture context). Since the conditional expectation of random effects given state sequences cannot be analytically derived, the proposed MCEM-like algorithm for semi-Markov switching linear mixed model cannot be transposed to the case of non-normally distributed observed data and other conditional restoration steps, for instance based on a Metropolis-Hastings algorithm, have to be derived for the random effects.

Acknowledgements

References

R.M. Altman. Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association*, 102:201–210, 2007.

G.E.B Archer and D.M. Titterington. Parameter estimation for hidden Markov chains. *Journal of Statistical Planning and Inference*, 108:365–390, 2002.

D. Barthélémy and Y. Caraglio. Plant architecture: a dynamic, multilevel and comprehensive approach of plant form and ontogeny. *Annals of Botany*, 99 (3):375–407, 2007.

B.S. Caffo, W. Jank, and G.L. Jones. Ascent-based Monte Carlo expectation-maximization. *Journal of the Royal Statistical Society, Series B*, 67(2):235–251, 2005.

S.G. Candy. Estimation in forest yield models using composite link functions with random effects. *Biometrics*, 53(1):146–160, 1997.

O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models.* Springer Series in Statistics. New York, NY: Springer. xvii, 652 p., 2005.

G. Celeux, O. Martin, and C. Lavergne. Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, 5:243–267, 2005.

F. Chaubert, Y. Caraglio, C. Lavergne, C. Trottier, and Y. Guédon. A statistical model for analyzing jointly growth phases, the influence of environmental factors and inter-individual heterogeneity. Applications to forest trees. In *5th International Workshop on Functional-Structural Plant Models*, pages P43, 1–3, Napier, New Zealand, nov 2007.

F. Chaubert-Pereira, Y. Caraglio, C. Lavergne, and Y. Guédon. Identifying ontogenetic, environmental and individual components of forest tree growth. *Submitted*, 2008.

S. Chib. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75:79–97, 1996.

R.G Cowell, A.P. Dawid, S.L. Lauritzen, and D.J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Science. New York, NY: Springer. xii, 321 p., 1999.

J.B. Durand, Y. Guédon, Y. Caraglio, and E. Costes. Analysis of the plant architecture via tree-structured statistical models: The hidden Markov tree models. *New Phytologist*, 166(3):813–825, 2005.

Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, 2002.

S. Fruhwirth-Schnatter. *Finite Mixture and Markov Switching models.* Springer Series in Statistics. New York, NY: Springer. xvii, 492 p., 2006.

Y. Guédon. Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12(3):604–639, 2003.

Y. Guédon. Exploring the state sequence space for hidden Markov and semi-Markov chains. *Computational Statistics and Data Analysis*, 51(5):2379–2409, 2007.

Y. Guédon, Y. Caraglio, P. Heuret, E. Lebarbier, and C. Meredieu. Analyzing growth components in trees. *Journal of Theoretical Biology*, 248(3):418–447, 2007.

M. Gupta, P. Qu, and J.G. Ibrahim. A temporal hidden Markov regression model for the analysis of gene regulatory networks. *Biostatistics*, 8(4):805–820, 2007.

S. Kim and P. Smyth. Segmental Hidden Markov Models with Random effects for Waveform Modeling. *Journal of Machine Learning Research*, 7:945–969, 2006.

C. Lavergne, M.J. Martinez, and C. Trottier. Finite mixture models for exponential repeated data. Technical Report 6119, INRIA, 2007.

G. Lindgren. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics, Theory and Applications*, 5: 81–91, 1978.

C.E. McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170, 1997.

G.J. McLachlan and T. Krishnan. *The EM algorithm and extensions. 2nd Edition*. Wiley Series in Probability and Statistics. New York, NY: John Wiley & Sons., 360 p., 2008.

X.L. Meng and S. Schilling. Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91(435):1254–1267, 1996.

R.M. Neal and G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In *Jordan, Michael I. (ed.), Learning in graphical models. Proceedings of the NATO ASI, Ettore Maiorana Centre, Erice, Italy, September 27 - October 7, 1996. Dordrecht: Kluwer Academic Publishers. NATO ASI Series. Series D. Behavioural and Social Sciences. 89, 355-368*. 1998.

Y. Ninomiya and A. Yoshimoto. Statistical Method for Detecting Structural Change in the Growth Process. *Biometrics*, 64(1):46–53, 2008. doi: 10.1111/j.1541-0420.2007.00844.x.

W. Qian and D.M. Titterington. Estimation of parameters in hidden Markov models. *Philosophical Transactions of the Royal Society, Series A*, 337(1647): 407–428, 1991.

F. Rijmen, E.H. Ip, S. Rapp, and E.G. Shaw. Qualitative longitudinal analysis of symptoms in patients with primary and metastatic brain tumours. *Journal of the Royal Statistical Society, Series A*, 171(3):739–753, 2008.

J.Q. Shi and S.Y. Lee. Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society, Series B*, 62(1): 77–87, 2000.

P. Smyth, D. Heckerman, and M.I. Jordan. Probabilistic independence networks for hidden Markov probability models. *Neural Computation*, 9(2):227–269, 1997.

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

Appendix: Forward-backward algorithm for sampling state sequences given random effects

Forward recursion

In the Markov switching linear mixed model case, the forward recursion is initialized for $t = 1$ by:

$$F_{aj}(1) = P\big(S_{a1} = j \mid Y_{a1} = y_{a1}, \xi_{a1}^J\big) \qquad j = 1, \ldots, J; \quad a = 1, \ldots, N;$$

$$= \frac{\phi(y_{a1}; X_{a1}\beta_j + \tau_j\xi_{aj}, \sigma_j^2)}{N_{a1}}\pi_j,$$

where $N_{a1} = P(Y_{a1} = y_{a1} \mid \xi_{a1}^J)$ is a normalizing factor with:

$$N_{a1} = \sum_{j=1}^{J} P(S_{a1} = j, Y_{a1} = y_{a1} \mid \xi_{a1}^J)$$

$$= \sum_{j=1}^{J} \phi(y_{a1}; X_{a1}\beta_j + \tau_j\xi_{aj}, \sigma_j^2)\,\pi_j.$$

For $t = 2, \ldots, T_a$, the forward recursion is given by:

$$F_{aj}(t) = P\big(S_{at} = j \mid Y_{a1}^t = y_{a1}^t, \xi_{a1}^J\big) \qquad j = 1, \ldots, J; \quad a = 1, \ldots, N;$$

$$= \frac{\phi(y_{at}; X_{at}\beta_j + \tau_j\xi_{aj}, \sigma_j^2)}{N_{at}} \sum_{i=1}^{J} p_{ij}F_{ai}(t-1).$$

The normalizing factor $N_{at} = P\big(Y_{at} = y_{at} \mid Y_{a1}^{t-1} = y_{a1}^{t-1}, \xi_{a1}^J\big)$ is obtained directly during the forward recursion as follows:

$$N_{at} = \sum_{j=1}^{J} P(S_{at} = j, Y_{at} = y_{at} \mid Y_{a1}^{t-1} = y_{a1}^{t-1}, \xi_{a1}^J)$$

$$= \sum_{j=1}^{J} \phi(y_{a1}; X_{a1}\beta_j + \tau_j\xi_{aj}, \sigma_j^2)\sum_{i=1}^{J} p_{ij}F_{ai}(t-1).$$

Backward pass

The backward pass can be seen as a stochastic backtracking procedure. The backward pass is initialized for $t = T_a$ by:

$$P\big(S_{aT_a} = j \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) = F_{aj}(T_a) \qquad j = 1, \ldots, J; \quad a = 1, \ldots, N.$$

The final state s_{aT_a} is sampled from the smoothed probabilities

$$\left(P\big(S_{aT_a} = j \mid Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big); j = 1, \ldots, J\right).$$

For $t = T_a - 1, \ldots, 1$, the backward pass is given by:

$$P\big(S_{at} = j \mid S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J\big) = \frac{p_{js_{a,t+1}}F_{aj}(t)}{\sum_{i=1}^{J} p_{is_{a,t+1}}F_{ai}(t)},$$

where the quantities $F_{aj}(t)$ are directly extracted from the forward recursion. The state s_{at} is sampled from the conditional distribution

$$\left(P\big(S_{at} = j \mid S_{a,t+1}^{T_a} = s_{a,t+1}^{T_a}, Y_{a1}^{T_a} = y_{a1}^{T_a}, \xi_{a1}^J \big); j = 1, \ldots, J \right).$$

Contents