Markov and semi-Markov switching linear mixed models used to identify forest tree growth components.

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

▶ To cite this version:

HAL Id: inria-00488100
https://hal.inria.fr/inria-00488100

Submitted on 12 Aug 2013

Markov and semi-Markov switching linear mixed models used to identify forest tree growth components

Florence Chaubert-Pereira[1],[*], Yann Guédon[1],[**], Christian Lavergne[2],[***] and Catherine Trottier[2],[****]

[1]CIRAD, UMR Développement et Amélioration des Plantes & INRIA, Virtual Plants,

TA A-96/02, 34398 Montpellier Cedex 5, France

[2]Montpellier 3 University, Institut de Mathématiques et de Modélisation de Montpellier,

Case Courrier 051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France

[*]*email:* chaubert@cirad.fr

[**]*email:* guedon@cirad.fr

[***]*email:* lavergne@math.univ-montp2.fr

[****]*email:* trottier@math.univ-montp2.fr

SUMMARY: Tree growth is assumed to be mainly the result of three components: (i) an endogenous component assumed to be structured as a succession of roughly stationary phases separated by marked change points that are asynchronous between individuals, (ii) a time-varying environmental component assumed to take the form of synchronous fluctuations between individuals, (iii) an individual component corresponding mainly to the local environment of each tree. In order to identify and 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 and their lengths (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 climatic 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: (i) sampling of state sequences given random effects, (ii) prediction of random effects given state sequences, (iii) maximization. The proposed statistical modeling approach is illustrated by the analysis of successive annual shoots along Corsican pine trunks influenced by climatic covariates.

1. Introduction

The analysis of plant structure at macroscopic scales is of major importance in forestry and different fields of agronomy; see Durand et al. (2005) and Guédon et al. (2007) for illustrations. Tree development can be reconstructed at a given observation date from morphological markers 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 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 that modulate plant development are mainly of climatic origin such as rainfall or temperature. The individual component may cover effects of diverse origins but always includes an individual genetic effect in the case of a tree population characterized by a genetic diversity. Other effects correspond to the local environment of each individual, e.g. competition between trees for light or nutrient resources.

A Gaussian hidden Markov chain (or Markov switching Gaussian model) (Ephraim and Merhav, 2002; Cappé et al., 2005) 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 $Y_t|S_t = s_t \sim \mathcal{N}(\mu_{s_t}, \Gamma_{s_t}^2)$. Markov switching linear models extend the class of Gaussian hidden Markov chains by incorporating the influence of covariates as fixed effects in the output process; see Fruhwirth-Schnatter (2006) for an overview of Markov switching models with different applications. 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 models and applied these models to longitudinal data; see also (Rijmen et al., 2008). Both Altman (2007) and Rijmen et al. (2008) assumed that the individual random effect and its variance are independent of the non-observable states.

A major drawback with hidden Markov models is the inflexibility in describing the time spent in a given state which is assumed to be geometrically distributed. It is unlikely that such a type of implicit state occupancy distribution is an appropriate model for tree growth phases. Guédon et al. (2007) applied hidden semi-Markov chains with simple observation distributions to forest tree growth data. Hidden semi-Markov chains (Guédon, 2003) generalize hidden Markov chains with the distinctive property of explicitly modeling the sojourn time in each state. Here we introduce semi-Markov switching linear mixed models that generalize both hidden semi-Markov chains and Markov switching linear mixed models. In our application context, the underlying semi-Markov chain represents 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- both the effect of time-varying climatic covariates as fixed effects and inter-individual heterogeneity as random effects. In this paper, we will consider two types of random effect: (i) an individual random effect (common across states), and (ii) individual state-specific random effects. In both cases, we assume that the random effect variance depends on the non-observable state. The objective is both to characterize the tree population and analyse the behavior of each individual within the population.

Since both the states of the underlying Markov (respectively semi-Markov) chain and the random effects are non-observable, Markov (respectively 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 sensitivity to starting values, slowness to converge and a strong computation burden. Since conditional independence assumptions within a Markov switching generalized linear mixed model can be represented by a directed acyclic graph, Rijmen et al. (2008) proposed to implement the E-step of the EM algorithm by a junction tree algorithm. 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 the random effects may be 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-specific random effects. Thus, the states are ordered and each state can be visited at most once. Their 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-specific 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).

For stochastic models involving a latent structure such as Markov or semi-Markov switching models, Markov chain Monte Carlo (MCMC) algorithms can be difficult to implement and suffer from problems such as poor mixing and the difficulty of diagnosing convergence; see Fearnhead (2008) and references therein. Here we propose a non-MCMC method which is a MCEM-like algorithm for estimating, in a general framework including as particular cases "left-right" and ergodic models, Markov and semi-Markov switching linear mixed models with either individual or individual state-specific random effects. Its iterations decompose into three steps: (i) sampling of state sequences given random effects, (ii) prediction of random effects given state sequences, (iii) maximization.

This paper is organized as follows. Markov and semi-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 using 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 Markovian model with finite state space $\{1, \ldots, J\}$. A 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 $\widetilde{p}_{ij} = P(S_t = j | S_{t-1} = i)$, $i, j = 1, \ldots, J$ with $\sum_j \widetilde{p}_{ij} = 1$.

A J-state semi-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

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

– absorbing state i: $\widetilde{p}_{ii} = P(S_t = i | S_{t-1} = i) = 1$ and for each $j \neq i$, $\widetilde{p}_{ij} = 0$.

- An explicit occupancy (or sojourn time) distribution is attached to each nonabsorbing state:

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

where U_j denotes the upper bound to the time spent in state j and $\sum_{u=1}^{U_j} d_j(u) = 1$.

We define as possible parametric state occupancy distributions binomial distributions, Poisson distributions and negative binomial distributions with an additional shift parameter h ($h \geqslant 1$) which defines the minimum sojourn time in a given state; see Web Appendix A for definitions of these parametric state occupancy distributions.

Since $t = 1$ is assumed to correspond to a state entering, the following relation is verified

$$P\left(S_{t+1} \neq j, S_{t-v} = j, v = 0, \ldots, t-1\right) = d_j(t)\,\pi_j.$$

Let $Y_{a,t}$ be the observation and $S_{a,t}$ be the non-observable state for individual a ($a = 1, \ldots, N$), at time t ($t = 1, \ldots, T_a$). Let $Y_{a,1}^{T_a} = (Y_{a,1}, Y_{a,2}, \ldots, Y_{a,T_a})$ denote the T_a-dimensional vector of observations for individual a, and $\boldsymbol{Y}$ the T-dimensional vector of all observations with $T = \sum_{a=1}^{N} T_a$. The vectors of non-observable states, $S_{a,1}^{T_a}$ and $\boldsymbol{S}$, are defined analogously.

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

- Individual random effect:

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

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

The individual status (compared to the average individual) within the population is common to all the states. The random effect ξ_a is thus common to all the states but the random effect standard deviation $\tau_{s_{a,t}}$ depends on the state.

- Individual state-specific random effect:

$$\text{Given state } S_{a,t} = s_{a,t}, \qquad Y_{a,t} = X_{a,t}\beta_{s_{a,t}} + \tau_{s_{a,t}}\xi_{a,s_{a,t}} + \epsilon_{a,t}, \qquad (2)$$

$$\xi_{a,s_{a,t}} \sim \mathcal{N}(0,1), \qquad \epsilon_{a,t} \,|\, S_{a,t} = s_{a,t} \sim \mathcal{N}(0, \sigma^2_{s_{a,t}}).$$

The individual status is different in each state. The random effect $\xi_{a,s_{a,t}}$ (with the attached standard deviation $\tau_{s_{a,t}}$) depends thus on the state.

In these definitions, $X_{a,t}$ is the Q-dimensional row vector of covariates for individual a at time t. Given the state $S_{a,t} = s_{a,t}$, $\beta_{s_{a,t}}$ is the Q-dimensional fixed effect parameter vector and $\sigma^2_{s_{a,t}}$ is the residual variance. The individuals are assumed to be independent.

In our context, the random effects ξ_a or $\xi_{a,s_{a,t}}$ are standardized (i.e. follow the standard Gaussian distribution). This formalization is in particular appropriate for the individual random effect model where the random effect variance changes with state; see Foulley and Quaas (1995). In the individual state-specific random effect models, the random effects for an individual a are assumed to be mutually independent ($\text{cov}(\xi_{a,j}, \xi_{a,j'}) = 0; j \neq j'$). Including random effects in the output process removes the assumption that the successive observations for an individual are conditionally independent given the non-observable states. In the individual state-specific random effect models, observations in different states for an individual a are assumed to be conditionally independent given states (for $t \neq t'$, $\text{cov}(Y_{a,t}, Y_{a,t'}|S^{T_a}_{a,1} = s^{T_a}_{a,1}) = 0$ if $s_{a,t} \neq s_{a,t'}$ and $\text{cov}(Y_{a,t}, Y_{a,t'}|S^{T_a}_{a,1} = s^{T_a}_{a,1}) = \tau^2_{s_{a,t}}$ if $s_{a,t} = s_{a,t'}$).

In the individual random effect models, observations in different states for an individual a are not assumed to be conditionally independent given states ($\text{cov}(Y_{a,t}, Y_{a,t'}|S_{a,1}^{T_a} = s_{a,1}^{T_a}) = \tau_{s_{a,t}}\tau_{s_{a,t'}}$ for $t \neq t'$). In state j, the introduction of random effects makes it possible to decompose the total variance $\Gamma_j^2 = \tau_j^2 + \sigma_j^2$ into two parts: variance due to inter-individual heterogeneity τ_j^2 and residual variance σ_j^2.

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

The maximum likelihood estimation is presented in the case of Markov switching linear mixed models with individual state-specific random effects. The extension to semi-Markov switching linear mixed models and the transposition to individual random effect models are straightforward. Markov switching linear mixed model parameters can be divided into two categories: parameters $\pi = (\pi_j; j = 1, \ldots, J)$ and $\widetilde{P} = (\widetilde{p}_{ij}; i, j = 1, \ldots, J)$ of the underlying Markov chain and 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 $\theta = (\pi, \widetilde{P}, \beta, \tau, \sigma^2)$ the set of parameters to be estimated.

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

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

where $\sum_{s_{a,1}^{T_a}}$ means "sum over every possible state sequences of length T_a for individual a".

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

the states $\boldsymbol{s}$ of the underlying Markov chain are observed:

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

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 cannot be applied because the successive observations for an individual are not conditionally independent given the non-observable states; see Section 2. The EM algorithm for a finite mixture of linear mixed models (Celeux et al., 2005) cannot be adapted because the distributions of $\xi_{a,j} \mid Y_{a,1}^{T_a} = y_{a,1}^{T_a}$ and $\xi_{a,j} \mid S_{a,t} = j, Y_{a,1}^{T_a} = y_{a,1}^{T_a}$ cannot be analytically derived for each individual a at time t. Thus, the Monte Carlo EM algorithm (Wei and Tanner, 1990), where the quantities computed in the deterministic E-step are approximated using Monte Carlo methods, provides an alternative approach.

For the presentation of the estimation algorithm, we adopted the framework of restoration-maximization (RM) algorithms proposed by Qian and Titterington (1991). The MCEM algorithm proposed by Altman (2007) can be interpreted as a RM algorithm with two restoration steps for the two latent structures, an 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 cannot adopt a similar approach with an unconditional restoration step since, in our definition of Markov switching linear mixed models (see Section 2), the random effects may be attached to the states. Hence, following Shi and Lee (2000), we chose rather to perform two conditional restoration steps, one for the state sequences given the random effects (and the observed data) and one for the random effects given the state sequences (and the observed data).

The proposed RM algorithm takes thus the following form:

1. Choose starting values $\theta^{(0)}$ and $\boldsymbol{\xi}^{(0)}$ for $k = 0$.

2. a. *Conditional restoration step (R-step) for state sequences*:

   For each individual a, sample M_k state sequences $s_{a,1}^{T_a}(m)$ from the conditional distribution $P(S_{a,1}^{T_a} = s_{a,1}^{T_a} \,|\, \xi_{a,1}^{J}(m), Y_{a,1}^{T_a} = y_{a,1}^{T_a}; \theta^{(k)})$ using:

   - a direct application of the forward-backward algorithm proposed by Chib (1996) in the Markov switching model case; see Web Appendix B;

   - a direct application of the forward-backward algorithm proposed by Guédon (2007) in the semi-Markov switching model case; see Web Appendix C.

   b. *Conditional restoration step for random effects; see Section 3.1*:

   For each individual a, compute for each sampled state sequence, $m = 1, \ldots, M_k$, the best posterior prediction

   $$\xi_{a,1}^{J}(m) = \mathrm{E}\Big(\xi_{a,1}^{J} \,|\, S_{a,1}^{T_a} = s_{a,1}^{T_a}(m), \; Y_{a,1}^{T_a} = y_{a,1}^{T_a}; \theta^{(k)}\Big).$$

   c. *Maximization step (M-step); see Section 3.2*:

   Choose $\theta^{(k+1)}$ that maximizes

   $$\frac{1}{M_k} \sum_{a=1}^{N} \sum_{m=1}^{M_k} \log f\Big(s_{a,1}^{T_a}(m), \xi_{a,1}^{J}(m), y_{a,1}^{T_a}; \theta^{(k)}\Big).$$

   d. *Sample size increase step*:

Set $k = k + 1$ and increase the number of sampled state sequences. If $M_{k+1} > M_k$, sample with replacement M_{k+1} random effects among the M_k predicted random effect.

3. If convergence is achieved, then declare $\theta^{(k+1)}$ to be maximum likelihood estimator; otherwise, return to step 2.

Since Markov (respectively semi-Markov) chain parameters and linear mixed model parameters form disjoint sets and influence the complete-data log-likelihood separately (see Equation 4 in the Markov switching model case and Web Equation 15 in Web Appendix D in the semi-Markov switching model case), Markov (respectively semi-Markov) chain parameters can be updated when the state sequences $\boldsymbol{S}$ are sampled, and linear mixed model parameters can be updated when the random effects $\boldsymbol{\xi}$ are predicted. It makes sense to re-estimate the parameters immediately before performing the conditional R-step for the other latent structure in order to speed up the convergence.

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 usual forward-backward algorithm and a backward pass for sampling state sequences; see Web Appendix B for Markov switching linear mixed models and Web Appendix C for semi-Markov switching linear mixed models.

3.1 *Prediction of random effects given state sequences*

The predicted vector for the random effects $\xi_{a,1}^J(m)$ attached to the mth state sequence sampled for individual a is:

$$
\begin{aligned}
\xi_{a,1}^J(m) &= \mathrm{E}\Big(\xi_{a,1}^J \mid S_{a,1}^{T_a} = s_{a,1}^{T_a}(m),\ Y_{a,1}^{T_a} = y_{a,1}^{T_a}; \theta^{(k)}\Big) \\
&= \Omega U_a'(m)\Big\{U_a(m)\Omega^2 U_a'(m) + \mathrm{diag}(U_a(m)\sigma^2)\Big\}^{-1}\Big(y_{a,1}^{T_a} - \sum_{j=1}^{J} I_{a,j}(m)X_a\beta_j\Big),\ (5)
\end{aligned}
$$

where

1 • $s_{a,1}^{T_a}(m)$ is the mth state sequence sampled for individual a,

2 • $\Omega = \mathrm{diag}(\tau_j; j = 1, \ldots, J)$ is the $J \times J$ random standard deviation matrix,

3 • $U_a(m)$ is the $T_a \times J$ design matrix associated with state sequence $s_{a,1}^{T_a}(m)$, composed of 1

4   and 0, where $u_{a,t}(m) = \Big(I(s_{a,t}(m) = 1), \ldots, I(s_{a,t}(m) = J)\Big)$ is the tth row with row sums,

5   $\sum_j U_a(m)(t,j) = 1$ and the sum of all elements, $\sum_t \sum_j U_a(m)(t,j) = T_a$,

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

7 • $\mathrm{diag}(U_a(m)\sigma^2)$ is the $T_a \times T_a$ diagonal matrix with $(u_{a,t}(m)\sigma^2; t = 1, \ldots, T_a)$ on its diagonal,

8 • $I_{a,j}(m) = \mathrm{diag}\Big\{I(s_{a,t}(m) = j), \ t = 1, \ldots, T_a\Big\}$ is a $T_a \times T_a$ diagonal matrix,

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

10 ## 3.2 *Maximization-step*

11 The M-step is presented in the case of Markov switching linear mixed models. The M-

12 step for semi-Markov switching linear mixed models is presented in Web Appendix D. In

13 the proposed MCEM-like algorithm, the conditional expectation of the complete-data log-

14 likelihood (Equation 4) given the observed data is approximated at iteration k by

$$\mathrm{E}\Big\{\log f(\boldsymbol{S}, \boldsymbol{\xi}, \boldsymbol{Y}; \theta) \,|\, \boldsymbol{Y} = \boldsymbol{y}; \theta^{(k)}\Big\} = \sum_{a=1}^{N} \mathrm{E}\Big\{\log f(S_{a,1}^{T_a}, \xi_{a,1}^{J}, Y_{a,1}^{T_a}; \theta) \,|\, Y_{a,1}^{T_a} = y_{a,1}^{T_a}; \theta^{(k)}\Big\}$$

$$\approx \ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k} \log f\Big(s_{a,1}^{T_a}(m), \xi_{a,1}^{J}(m), y_{a,1}^{T_a}; \theta^{(k)}\Big)$$

$$\approx \ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{j=1}^{J} I\Big(s_{a,1}(m) = j\Big)\log \pi_j^{(k)}$$

$$+ \ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{t=2}^{T_a}\sum_{i,j=1}^{J} I\Big(s_{a,t}(m) = j, \ s_{a,t-1}(m) = i\Big)\log \widetilde{p}_{ij}^{(k)}$$

$$+ \ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{j=1}^{J} \log \phi(\xi_{a,j}(m); 0, 1)$$

$$+ \ \frac{1}{M_k}\sum_{a=1}^{N}\sum_{m=1}^{M_k}\sum_{t=1}^{T_a}\sum_{j=1}^{J} I\Big(s_{a,t}(m) = j\Big)\log \phi(y_{a,t}; X_{a,t}\beta_j^{(k)} + \tau_j^{(k)}\xi_{a,j}(m), \sigma_j^{2(k)}). \tag{6}$$

At iteration k, the new values for the parameters of the Markov switching linear mixed model are obtained by maximizing the different terms of Equation 6, each term depending on a given subset of θ.

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

- initial probabilities

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

- transition probabilities

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

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

- fixed effect parameters

$$\beta_j^{(k+1)} = \left(\sum_a \sum_m X_a' I_{a,j}(m) X_a \right)^{-1} \left\{ \sum_a \sum_m X_a' I_{a,j}(m) \Big(y_{a,1}^{T_a} - \tau_j^{(k)} \xi_{a,j}(m)\Big) \right\}, \tag{7}$$

- random effect standard deviations

$$\tau_j^{(k+1)} = \frac{\sum_a \sum_m \sum_t I\Big(s_{a,t}(m) = j\Big) \xi_{a,j}(m) \Big(y_{a,t} - X_{a,t}\beta_j^{(k)}\Big)}{\sum_a \sum_m \sum_t I\Big(s_{a,t}(m) = j\Big) \xi_{a,j}^2(m)}, \tag{8}$$

- residual variances

$$\sigma_j^{2(k+1)} = \frac{\sum_a \sum_m \Big(y_{a,1}^{T_a} - X_a\beta_j^{(k)} - \tau_j^{(k)}\xi_{a,j}(m)\Big)' I_{a,j}(m) \Big(y_{a,1}^{T_a} - X_a\beta_j^{(k)} - \tau_j^{(k)}\xi_{a,j}(m)\Big)}{\sum_a \sum_m \text{tr}(I_{a,j}(m))}. \tag{9}$$

These reestimation formulas are thus similar to standard maximum likelihood estimators for the different parameters.

For the computation of standard errors, Wei and Tanner (1990) proposed a Monte Carlo version of Louis method where the expectation is replaced by a Monte Carlo procedure when using a MCEM algorithm. The Wei and Tanner method can directly be applied in the proposed MCEM-like algorithm.

1. 3.3 *MCEM-like algorithm for individual random effect models*

2. The application to individual random effect models (linear mixed model (1)) is straightfor-

3. ward. Since the individual random effects are incorporated in the output process, the main

4. difference concerns the conditional R-step of random effect prediction given a state sequence

5. (Equation 5). Using the notations introduced in Section 3.1, the predicted random effect ξ_a

6. for each individual a is given by

$$
\begin{aligned}
\xi_a(m) &= \mathrm{E}\Big(\xi_a \mid S_{a,1}^{T_a} = s_{a,1}^{T_a}(m), Y_{a,1}^{T_a} = y_{a,1}^{T_a}; \theta^{(k)}\Big) \\
&= \tau' U_a'(m)\Big\{U_a(m)\tau\tau' U_a'(m) + \mathrm{diag}(U_a(m)\sigma^2)\Big\}^{-1}\Big(y_{a,1}^{T_a} - \sum_{j=1}^{J} I_{a,j}(m)X_a\beta_j\Big),
\end{aligned}
$$

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

8. forward-backward algorithm (see Web Appendix B for Markov switching linear mixed models

9. and Web Appendix C for semi-Markov switching linear mixed models) and in the M-step

10. (Equations 7, 8 and 9), the random effects $\xi_{a,1}^J$ are replaced by ξ_a.

11. 3.4 *Initialisation of the algorithm*

12. Various simulations were conducted using different starting values. The more distant the

13. starting values from true values, the worse the parameter estimates. We recommend choosing

14. as starting values the parameters estimated by the EM algorithm for a simple Markov or

15. semi-Markov switching linear model (i.e. without random effects).

16. 3.5 *Convergence of the algorithm*

17. Under the assumption of convergence of random effect predictions, we chose to monitor

18. the convergence of the proposed MCEM-like algorithm by the difference between successive

19. iterations

$$
\log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}^{(k+1)}; \theta^{(k+1)}) - \log P(\boldsymbol{Y} = \boldsymbol{y} \mid \boldsymbol{\xi}^{(k)}; \theta^{(k)}). \tag{10}
$$

1 The quantity $\log P(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{\xi}^{(k)}; \theta^{(k)})$ is obtained directly as a byproduct of the forward

2 recursion.

3 *3.6 Sample size*

4 The conditional R-steps rely on the restoration of several pairs $(s_{a,1}^{T_a}, \xi_{a,1}^J)$ for each individual

5 a. As discussed by Wei and Tanner (1990), it is inefficient to start with a large number of

6 sampled state sequences M_k. They recommended increasing M_k as the current approximation

7 moves closer to the true maximizer.

8 **4. Application to Corsican pine growth**

9 The use of semi-Markov switching linear mixed models is illustrated here by an analysis

10 of forest tree growth. The data set comprised four sub-samples of Corsican pines planted

11 in a forest stand in the "Centre" region (France): 31 6-year-old trees, 29 12-year-old trees

12 (first year not measured), 30 18-year-old trees (first year not measured) and 13 23-year-old

13 trees (three first years not measured). Tree trunks were described by annual shoot from the

14 top to the base where the length (in cm) was recorded for each annual shoot (Figure 1).

15 The annual shoot is defined as the segment of stem established within a year. The trees

16 were chosen in order to cover the entire range of behaviors and were not subject to any

17 silvicultural interventions.

18 [Figure 1 about here.]

19 We applied the practical approach discussed in Guédon et al. (2007) for determining the

20 number of growth phases i.e. the number of states of the underlying non-ergodic semi-Markov

21 chain. A "left-right" three-state semi-Markov switching linear mixed model composed of two

22 successive transient states followed by a final absorbing state was estimated. Since the last

23 year of measurement is arbitrary with regard to tree development, the length of the last

growth phase is assumed to be systematically right-censored and cannot be modeled. 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 or elongation. We chose to use an intercept and the centered cumulative 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:

$$\text{Given state } S_{a,t} = j, \qquad y_{a,t} = \beta_{j1} + \beta_{j2}X_t + \tau_j\xi_{a,j} + \epsilon_{a,t},$$

$$\xi_{a,j} \sim \mathcal{N}(0,1), \qquad \epsilon_{a,t} \,|\, S_{a,t} = j \sim \mathcal{N}(0,\sigma_j^2),$$

where $y_{a,t}$ is the length of the annual shoot for individual a at time t, β_{j1} is the intercept, X_t is the centered cumulative rainfall at time t ($\mathrm{E}(X_t) = 0$), β_{j2} is the cumulative rainfall parameter and $\xi_{a,j}$ is the random effect for individual a in state j. Because of the centering of the climatic covariate, the intercept β_{j1} is directly interpretable as the average length of successive annual shoots in state j.

[Figure 2 about here.]

The MCEM-like algorithm was initialized with parameters π, P, d, β and σ^2 estimated without taking the random effects into account (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 10). The plot of the values against the iteration rank showed that the estimation algorithm converged rapidly by, say, about 70 iterations with $M_k = k$ state sequences sampled for each tree at the kth iteration (Figure 2). It should be noted that various strategies were tested for increasing the number of sampled state sequences: linearly (McCulloch, 1994), in stages (McCulloch, 1997) and geometrically (Caffo et al., 2005). For the Corsican pine data set, the parameter estimates were not sensitive to the selected strategy.

1   The MCEM-like algorithm was initialized with a "left-right" model such that $\pi_j > 0$ for

2 each state j, $p_{ij} = 0$ for $j \leqslant i$ and $p_{ij} > 0$ for $j > i$ for each transient state i. The fact

3 that states 1 and 2 are the only possible initial states (with $\pi_1 = 0.95$ and $\pi_2 = 0.05$ at

4 convergence) and that state 2 cannot be skipped (i.e. $p_{13} = 0$ at convergence) is the result of

5 the iterative estimation procedure; see Figure 3. The estimated transition probability matrix

6 is thus degenerate i.e. for each transient state i, $p_{i\,i+1} = 1$ and $p_{ij} = 0$ for $j \neq i+1$. This

7 deterministic succession of states supports the assumption of a succession of growth phases.

8 It should be noted that the estimated state occupancy distributions for states 1 and 2 are

9 far from geometric distributions in terms of shape and relative dispersion; see Figure 3. This

10 is an *a posteriori* justification of the semi-Markovian modeling of the growth phases.

11   The state occupancy distributions estimated for a Gaussian hidden semi-Markov chain

12 (GHSMC) (i.e. without taking into account the effect of climatic covariates and inter-

13 individual heterogeneity) and for a semi-Markov switching linear mixed model (SMS-LMM)

14 were compared. When the effect of climatic covariates and inter-individual heterogeneity were

15 taken into account, this rendered the growth phases more synchronous between individuals;

16 see the estimated state occupancy distributions, in particular their standard deviations, for

17 the two models in Table 1 and Figure 3.

18   [Figure 3 about here.]

19   At convergence of the MCEM-like algorithm, the median predicted random effects were

20 computed for each individual based on the random effects predicted for each state in each

21 sampled state sequence. The most probable state sequence given the median predicted

22 random effects was computed for each observed sequence using a Viterbi-like algorithm

23 (Guédon, 2003). This restored state sequence can be viewed as the optimal segmentation of

24 the corresponding observed sequence into sub-sequences, each corresponding to a given state.

25 The optimal segmentations of the observed sequences were used, in particular, to compute

the mean centered cumulative rainfall and the average cumulative rainfall effect in each state

(see below).

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 cumulative rainfall in state j. The marginal observation distribution represents the length of the annual shoots in state j. The marginal observation distributions for the different states are well separated (little overlap 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. The fixed part of the three 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 overlap and correspond to a growth increase.

[Table 1 about here.]

The average cumulative rainfall effect (i.e. the average amplitude of the climatic fluctuations) was computed as $\beta_{j2} \times \mathrm{mad}_j(X)$ for each state j where $\mathrm{mad}_j(X)$ is the mean absolute deviation of the centered cumulative rainfall in state j. The effect of cumulative rainfall was weak in the first state (of slowest growth) while it was stronger 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 approximately 55%) and decreased slightly in the last state (approximately 47%).

[Figure 4 about here.]

The correlation coefficient between the predicted random effect in state 1 and the predicted random effect in state 2 was 0.28 while the correlation coefficient between the predicted random effect in state 2 and the predicted random effect in state 3 was 0.61. Hence, the

behavior of an individual is quite strongly related between the last two states but only loosely related between the first two states. A 95% prediction interval (Hulting and Harville, 1991) was computed to check whether the influence of the predicted random effect for each state is significant:

$$[-t_{0.975}(N-1)\frac{\sigma_{\xi_j}}{\sqrt{N}};\ t_{0.975}(N-1)\frac{\sigma_{\xi_j}}{\sqrt{N}}]$$

where σ_{ξ_j} is the empirical standard deviation of the predicted random effects for state j. Of the 103 Corsican pines, 50 had a significant random effect in each state and of these 50 individuals, 29 showed a common behavior over all growth phases (i.e. growing either more rapidly or slowly than the "average tree" in all growth phases) but 21 showed different behaviors between growth phases. Hence, the more general assumption of individual state-specific random effect models (linear mixed model (2)) compared to individual random effect models (linear mixed model (1)) is more representative of Corsican pine behavior.

Complementary biological results concerning Corsican pine and sessile oak growth can be found in Chaubert-Pereira et al. (2009).

5. Concluding remarks

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).

The estimation algorithms proposed in this paper can be directly 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 categorical or count response variables (for instance, apex

death/life, non-flowering/flowering character, number of growth units per annual shoot in the plant architecture context). Since the conditional expectation of random effects given state sequences cannot be derived analytically, the proposed MCEM-like algorithm for semi-Markov switching linear mixed models cannot be transposed to semi-Markov switching generalized linear mixed models, and other conditional restoration steps, for instance based on a Metropolis-Hastings algorithm, have to be derived for the random effects.

Acknowledgements

References

Altman, R. (2007). 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.

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

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

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

Chaubert-Pereira, F., Caraglio, Y., Lavergne, C., and Guédon, Y. (2009). Identifying

ontogenetic, environmental and individual components of forest tree growth. *Annals of Botany, in press* .

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

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

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

Fearnhead, P. (2008). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. *Statistics and Computing* **18,** 151–171.

Foulley, J. and Quaas, R. (1995). Heterogeneous variances in Gaussian linear mixed models. *Genetics Selection Evolution* **27,** 211–228.

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

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

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

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

Hulting, F. and Harville, D. (1991). Some Bayesian and non-Bayesian procedures for the analysis of comparative experiments and for small area estimation: Computational aspects, frequentist properties, and relationships. *Journal of the American Statistical Association* **86,** 557–568.

Kim, S. and Smyth, P. (2006). Segmental hidden Markov models with random effects for waveform modeling. *Journal of Machine Learning Research* **7,** 945–969.

McCulloch, C. (1994). Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association* **89,** 330–335.

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

McLachlan, G. and Krishnan, T. (2008). *The EM algorithm and extensions. 2nd Edition.* Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons., 360 p.

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

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

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

Wei, G. and Tanner, M. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* **85,** 699–704.
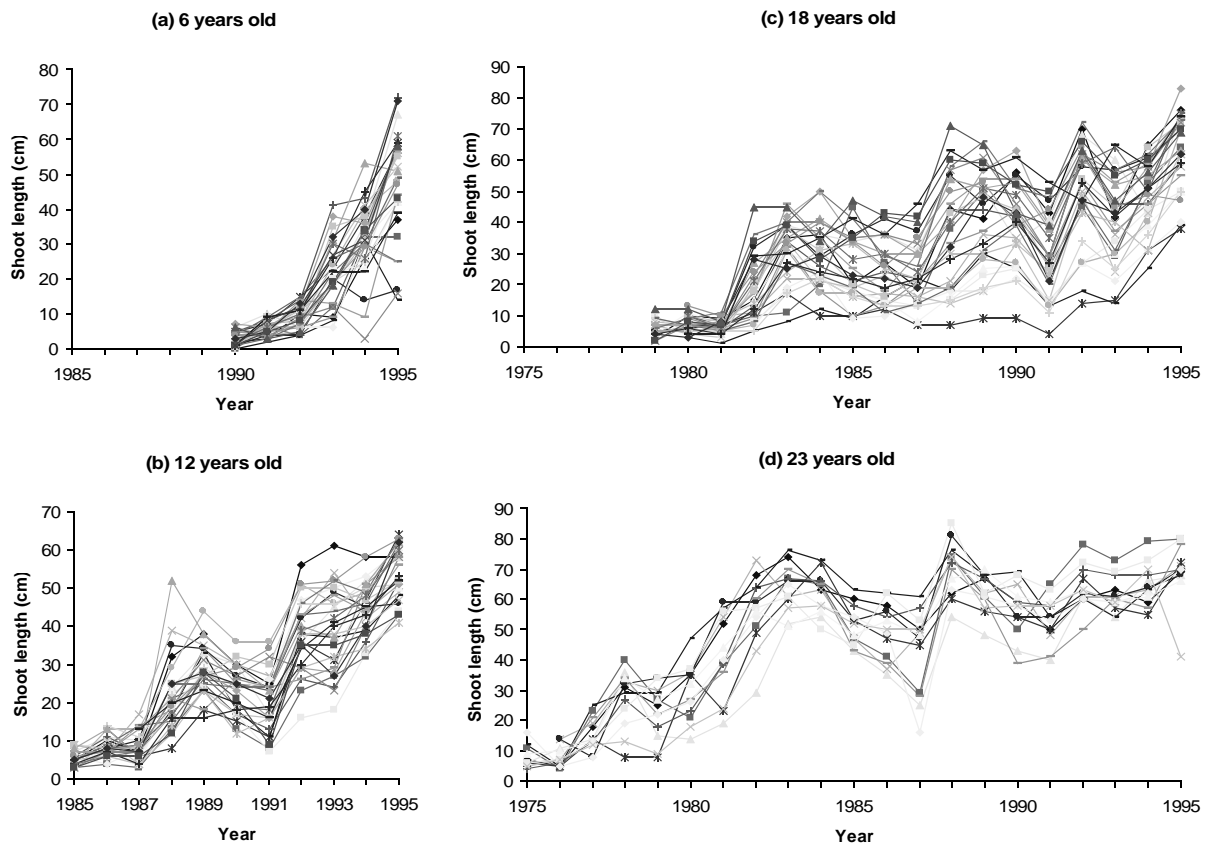
Figure 1. Length of successive annual shoots along Corsican pine trunks: (a) 6-year-old trees, (b) 12-year-old trees, (c) 18-year-old trees, (d) 23-year-old trees.
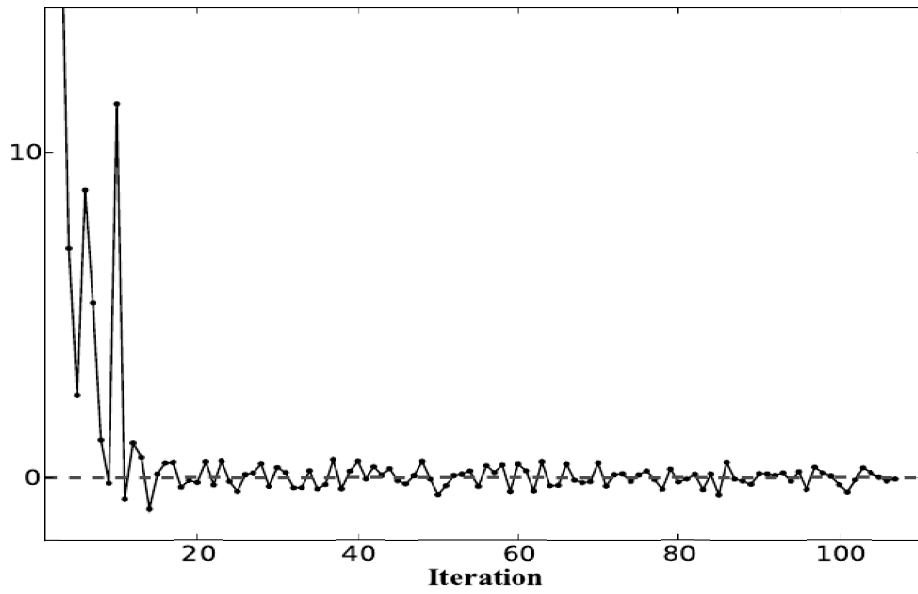
Figure 2. Difference of observed data log-likelihood given random effects between successive iterations (Equation 10) for the estimated semi-Markov switching linear mixed model with individual state-specific random effects.
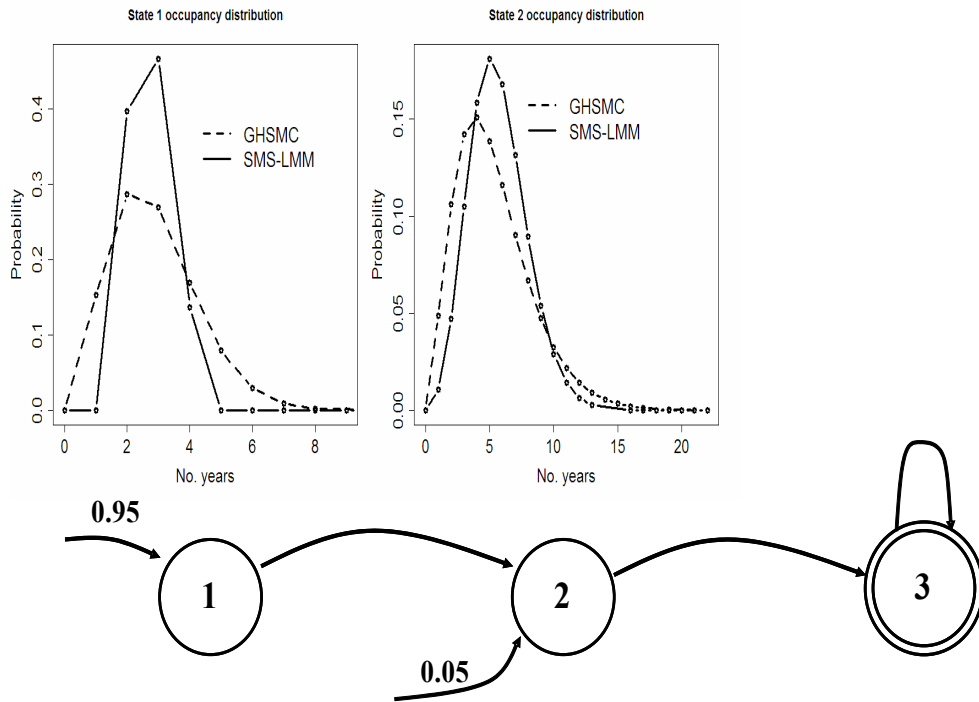
Figure 3. Estimated underlying semi-Markov chain. Each state is represented by a vertex which is numbered. Vertices representing transient states are edged by a single line while the vertex representing the final absorbing state is edged by a double line. Possible transitions between states are represented 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 for a Gaussian hidden semi-Markov chain (GHSMC) and the point lines correspond to occupancy distributions estimated for a semi-Markov switching linear mixed model (SMS-LMM).
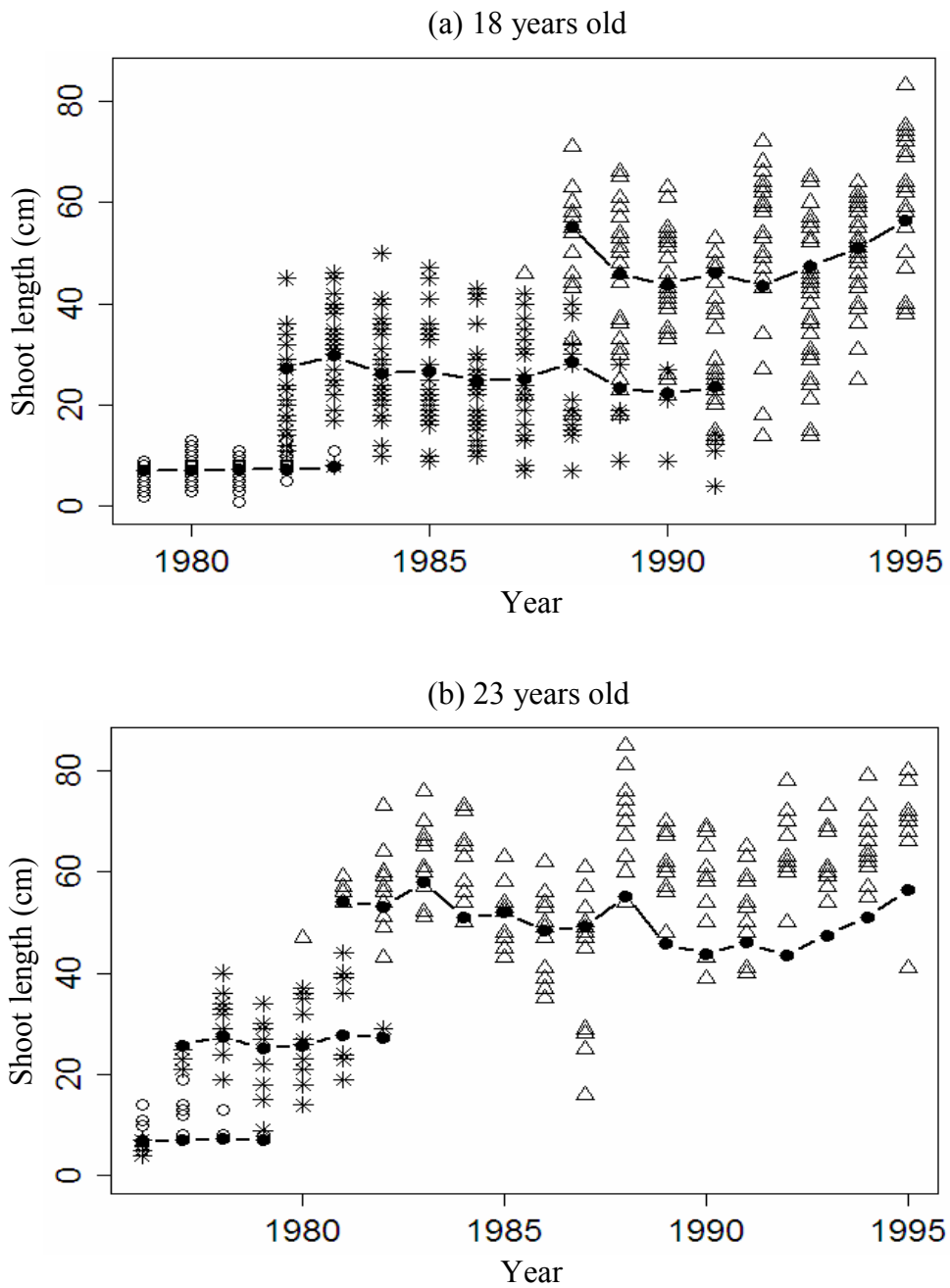
(a) 18 years old

(b) 23 years old

Figure 4. Fixed part of the three observation linear mixed models (i.e. $\beta_{j1} + \beta_{j2} X_t$ for each state j; ○: state 1, ∗: state 2, △: state 3) represented by point lines and observed annual shoot lengths (points): (a) 18-year-old trees, (b) 23-year-old trees.

		State j		
		1	2	3
Occupancy distributions (year) mean, sd	GHSMC	$P(1, 1.88)$ $2.88, 1.37$	$NB(1, 4.36, 0.5)$ $5.31, 2.93$	
	SMS-LMM	$B(2, 4, 0.37)$ $2.73, 0.68$	$NB(1, 73.29, 0.94)$ $5.56, 2.20$	
Regression parameters (SMS-LMM)	Intercept β_{j1} (cm) (s.e.)	7.09 (0.14)	25.79 (0.3)	50.25 (0.39)
	Cumulative rainfall parameter β_{j2} cm.mm^{-1} (s.e.)	$2.7 \ 10^{-3}$ $(1.2 \ 10^{-3})$	$16.5 \ 10^{-3}$ $(2 \ 10^{-3})$	$30.9 \ 10^{-3}$ $(2.4 \ 10^{-3})$
	Average cumulative rainfall effect $\beta_{j2} \times \mathrm{mad}_j(X)$ (cm)	0.23	1.71	3.76
Variability decomposition (SMS-LMM)	Random variance τ_j^2 (s.e.)	5.79 (2.23)	49.89 (2.14)	69.39 (2.37)
	Residual variance σ_j^2 (s.e.)	4.74 (0.38)	39.95 (2.48)	76.86 (4.61)
	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 (cm) 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 (i.e. where the influence of covariates and the inter-individual heterogeneity are not taken into account) with the estimated semi-Markov switching linear mixed model (SMS-LMM) parameters (state occupancy distributions and marginal observation distributions). The regression parameters, the cumulative rainfall effect and the variability decomposition are given for each observation linear mixed model. Standard errors are given in brackets.