A statistical model for analyzing jointly growth phases, the influence of environmental factors and inter-individual heterogeneity. Applications to forest trees.

Florence Chaubert[1], Yves Caraglio[2], Christian Lavergne[3], Catherine Trottier[3] & Yann Guédon[1]

[1] *CIRAD, UMR Développement et Amélioration des Plantes*
INRIA, Equipe de recherche Virtual Plants
Avenue Agropolis, TA A-96/02, 34398 Montpellier Cedex 5, France
E-mail: chaubert@cirad.fr, guedon@cirad.fr

[2] *CIRAD, UMR botAnique et bioinforMatique de l'Architecture des Plantes*
TA A-51/PS2, 34398 Montpellier Cedex 5, France
E-mail: caraglio@cirad.fr

[3] *UMR Institut de Mathématiques et de Modélisation de Montpellier,*
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
E-mail: trottier@math.univ-montp2.fr, Christian.Lavergne@math.univ-montp2.fr

Keywords: Markov switching model, linear mixed model, growth phases, ontogeny.

Introduction

Observed growth, as given for instance by the length of successive annual shoots along a tree trunk, is mainly the result of two components: an ontogenetic component and an environmental component. The ontogenetic component is assumed to be structured as a succession of roughly stationary phases, asynchronous between individuals, while the time-varying environmental component (mainly of climatic origin) is assumed to take the form of local fluctuations, synchronous between individuals. Guédon *et al.* (2007) proposed a set of methods for analyzing these components. In their approach, the environmental component was simply characterized globally but it may be interesting to study in detail the plant's response to changing climatic conditions. Here, we propose a statistical modeling approach that enables to differentiate the role of environmental explanatory variables and the part of inter-individual heterogeneity between the growth phases.

First results for a sample of 18-year-old Corsican pines and another sample of 15-year-old sessile oaks were reported in Vera *et al.* (2004). We propose to extend these analyses to data samples grouping for each species (Corsican pine and sessile oak) several sub-samples corresponding to different age classes, all the trees having grown during a common range of years. In this way, different climatic years are mixed for a given ontogenetic phase.

Markov switching linear mixed model

Vera *et al.* (2004) introduced Markov switching linear mixed models, i.e. models that combine linear mixed models (Verbeke and Molenberghs, 2000) in a markovian manner. These models belong to the family of hidden Markovian models (Ephraim and Merhav, 2002). The underlying Markov chain represents the succession of growth phases while the linear mixed models attached to each state of the Markov chain represent both the influence of time-varying climatic explanatory variables and inter-individual heterogeneity due to unobserved factors such as plant pathogen infestation within a given growth phase. The influence of climatic explanatory variables is modeled as a fixed effect and the inter-individual heterogeneity as a random effect (Verbeke and Molenberghs, 2000). Thus, the introduction of random effects makes it possible to decompose the total variability into two parts: variability due to inter-individual heterogeneity and residual variability.

In temperate regions, rainfall can have a one-year-delayed effect (on the number of elements) or an immediate effect (on shoot elongation) depending on whether it occurs during the organogenesis or the elongation. We therefore chose to use rainfall cumulated during a selected period

of the previous year (organogenesis) and rainfall cumulated during a selected period of the current year (elongation) as climatic explanatory variables. These explanatory variables may be considered quite rough compared to explanatory variables computed on the basis of sophisticated ecophysiological models.

We propose two nested families of Markov switching linear mixed models which differ in the assumptions made concerning inter-individual heterogeneity:

- model 1: The individual tree status within the population (tree growing either quickly or slowly than the average tree) is common to all the growth phases;
- model 2: The individual tree status is different in each growth phase.

Markov switching linear mixed models were estimated using a Stochastic or Monte-Carlo Expectation Maximization (SEM or MCEM) algorithm (McLachlan and Krishnan, 1997) whose elementary iteration decomposes into three steps: sampling of state sequences, random effect prediction and maximization.

We are interested both by the quantification of the inter-individual heterogeneity of the tree population and by the behavior of each individual within the population.

Applications to forest trees

Sessile oaks

The data sample comprised two sub-samples: 46 15-year-old trees (from 1983 to 1997) and 20 29-year-old trees (24 last years measured from 1974 to 1997). These trees originating from natural regenerations were observed in a private forest near Louppy-le-château (north-east France). It should be noted that the silvicultural practices favored synchronous germinations in the years following mass fruiting . Stand density was 2000 stems/ha. Tree trunks were described by annual shoot where the length (in cm) was recorded for each annual shoot.

Length of annual shoots

"Left-right" two-state Markov switching linear mixed models composed of a transient state followed by a final absorbing state were estimated on the basis of the two sub-samples of sessile oaks. The inter-individual heterogeneity is greater at the beginning of the plant life (first growth phase) while the influence of climatic explanatory variables is roughly proportional to growth level. For the two models, the part of variability due to inter-individual heterogeneity is less than 8% in the second phase (of strongest growth). Moreover, the more general assumption of model 2 (a random effect attached to each state) seems more realistic. For example, a tree can grow more slowly in the first phase than the average tree and then more quickly in the second phase. The small part of variability due to inter-individual heterogeneity can be explained by the fact that the individuals were selected among the dominant or codominant trees, by the thinning and by the synchronous natural regeneration for each sub-sample.

Corsican pines

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). Plantation density was 1800 stems/ha for the first sub-sample (6 year old) and 2200 stems/ha for the three other sub-samples. Tree trunks were described by annual shoot where two quantitative variables were recorded for each annual shoot, namely length (in cm) and number of branches per tier. These trees were not subject to any silvicultural interventions.

Length of annual shoots

"Left-right" three-state Markov switching linear mixed models composed of two successive transient states followed by a final absorbing state were estimated. As for the sessile oaks, model 2 seems more realistic although the behaviors on the last two phases (of strongest growths) are generally similar. On the first two growth phases, climatic explanatory variables have a moderate influence on the length of annual shoots. On the phase of strongest growth, the annual shoots

length is markedly influenced by the current year rainfall (elongation) and a little less by the previous year rainfall (organogenesis). As for sessile oaks, the inter-individual heterogeneity is greater at the beginning of the plant life (cf Table 1). The important heterogeneity can be explained by the time spent in nursery, the lack of thinning and the sampling strategy (trees chosen in order to cover the whole range of behaviors).

	Intercept	Current year rainfall (Elongation)	Previous year rainfall (Organogenesis)	Part of inter-individual heterogeneity	Residual variability
Growth phase 1	4.69	6.62×10^{-3}	-4.43×10^{-3}	64.81%	10.16
Growth phase 2	22.4	2.89×10^{-3}	2.54×10^{-3}	65.98%	95.13
Growth phase 3	33.04	$\mathbf{25.31} \times 10^{-3}$	$\mathbf{10.32} \times 10^{-3}$	49.62%	166.19

Table 1: Corsican pines, Model 2 - linear mixed model estimated on each growth phase.

Number of branches per tier

It should be noted that the number of branches per tier is strongly correlated with the annual shoot length ($r = 0.66$). Hence, when introducing the annual shoot length as an explanatory variable, it is no more necessary to model explicitly the phases for the "number of branches per tier" response variable. A simple linear mixed model was thus estimated where both the annual shoot length, the previous year rainfall (annual shoot organogenesis) and the current year rainfall (annual shoot elongation and branch organogenesis) were taken into account as explanatory variables (and their influences modeled as fixed effects). The influence of the previous year rainfall and the current year rainfall is not significant while as expected, the influence of the annual shoot length is strong. Since most of the inter-individual heterogeneity is already reflected in the "length of the annual shoot" explanatory variable, the part of variability due to the supplementary inter-individual heterogeneity is less than 14%.

Discussion

We note the same behavior for these species: the inter-individual heterogeneity is greater at the beginning of the plant life (first growth phases) while the influence of climatic explanatory variables is roughly proportional to growth level. However, the tree origin, the silvicultural interventions and the sampling strategy seems to influence the inter-individual heterogeneity part. To understand which among these three factors have the greatest influence is a challenging problem. It would be very useful to study the part of inter-individual heterogeneity for various species in similar conditions or for a given species in various conditions.

Another approach to take into account the influence of the environment on plant growth would be to introduce a year random effect common to all the individuals to model the synchronous part of the inter-annual fluctuations. The total variability would decompose into three parts: variability due to the environment (year random effect), variability due to the inter-individual heterogeneity (individual random effect) and residual variability.

References

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

Guédon, Y., Caraglio, Y., Heuret P., Lebarbier, E., and Meredieu, C. (2007). Analyzing growth components in tree. To be published in *Journal of Theoretical Biology*.

McLachlan, G.J. and Krishnan, T. (1997) *The EM algorithm and extensions*, New York: Wiley.

Véra, C., Guédon, Y., Lavergne, C. and Caraglio, Y. (2004). Analysis of longitudinal data applied to plant architecture study In: 4*th International Workshop on Functional-Structural Plant Models*, Montpellier, 65-69.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.