

1 **Deciphering the developmental plasticity of walnut saplings in**
2 **relation to climatic factors and light environment**

3
4 **Olivier Taugourdeau¹, Florence Chaubert-Pereira², Sylvie Sabatier¹, Yann Guédon^{2*}**

5 ¹CIRAD/UM2, UMR botAnique et bioinforMatique de l'Architecture des Plantes, TA A-
6 51/PS2, 34398 Montpellier Cedex 5, France (olivier.taugourdeau@cirad.fr, sylvie-
7 annabel.sabatier@cirad.fr)

8 ²CIRAD, UMR Amélioration Génétique et Adaptation des Plantes méditerranéennes et
9 tropicales and INRIA, Virtual Plants, TA A-96/02, 34398 Montpellier Cedex 5, France
10 (chauchau64@gmail.com, guedon@cirad.fr)

11
12 *Corresponding author: guedon@cirad.fr,

13
14 5 Tables, 7 Figures

15 Running title: Developmental plasticity of walnuts saplings

16
17 **Abstract**

18 Developmental plasticity, the acclimation of plants to their local environment, is known to be
19 crucial for the fitness of perennial organisms such as trees. However, deciphering the many
20 possible developmental and environmental influences involved in such plasticity in natural
21 conditions requires dedicated statistical models integrating developmental phases,
22 environmental factors and inter-individual heterogeneity. These models should be able to
23 analyze retrospective data (number of leaves or length of annual shoots along the main stem in
24 our case). In this study Markov switching linear mixed models were applied to the analysis of
25 the developmental plasticity of walnut saplings during the establishment phase in a mixed

1 Mediterranean forest. In the Markov switching linear mixed models estimated from walnut data
2 sets, the underlying Markov chain represents both the succession and lengths of growth phases,
3 while the linear mixed models represent both the influence of climatic factors and inter-
4 individual heterogeneity within each growth phase. On the basis of these integrative statistical
5 models, it is shown that walnut saplings have an opportunistic mode of development that is
6 primarily driven by the changing light environment. In particular, light availability explains the
7 ability of a tree to reach a phase of strong growth where the first branches can appear. It is also
8 shown that growth fluctuation amplitudes in response to climatic factors increased while inter-
9 individual heterogeneity decreased along tree development.

10

11 Key words: growth components; *Juglans regia* L.; linear mixed model; Markov switching
12 model; ontogeny; plant architecture.

13

14 **Introduction**

15 In natural conditions the developmental plasticity of trees is one of the main determinants of
16 forest successional dynamics. In particular, the early life period to a large extent conditions the
17 modalities of replacement of senescent trees by young trees. In understory, tree growth is
18 driven by light availability with substantial differences according to shade tolerance (Wright *et*
19 *al.*, 2000) and climatic factors (Sanchez-Gomez *et al.*, 2008). Here, our objective was to study
20 the establishment phase of trees by focusing on apical growth and branching of young trees
21 growing in a mixed forest. One of the major difficulties when studying developmental plasticity
22 is the “real-world complexity of multiple biotic and abiotic variables” (Miner *et al.*, 2005).
23 Hence a decomposition approach should be applied to identify and characterize the different
24 variables that influence tree development.

25

1 Tree structure development can be reconstructed at a given observation date from external
2 morphological markers (such as cataphyll and branching scars) that correspond to past events
3 (Barthélémy and Caraglio, 2007). Observed apical growth, as given for instance by the length
4 of successive annual shoots along a tree main stem, is assumed to result mainly from three
5 components: an ontogenetic component, an environmental component and an individual
6 component (Chaubert-Pereira *et al.*, 2009). The ontogenetic component is assumed to be
7 structured as a succession of roughly stationary growth phases that are asynchronous between
8 individuals (Guédon *et al.*, 2007). A phase is said to be stationary if there is no systematic
9 change in mean (no trend), no systematic change in variance and no strictly periodic variation
10 of the variable of interest. The key question tackled in Guédon *et al.* (2007) was whether the
11 ontogenetic growth component along an axis at the growth unit or annual shoot scale takes the
12 form of a trend (i.e. a gradual change in the mean level) or of a succession of phases. Their
13 results support the assumption of abrupt changes between roughly stationary phases rather than
14 gradual changes. The environmental component is assumed to take the form of local
15 fluctuations that are synchronous between individuals. This environmental component is thus
16 assumed to be a “population” component as opposed to the individual component which
17 represents the growth level deviation in each phase of a tree with reference to the “average” tree.
18 The individual component may cover effects of diverse origins but always includes a genetic
19 effect (Sabatier *et al.*, 2003a; Segura *et al.*, 2008). Other effects correspond to the tree's local
20 environment (e.g. light resources); see Pinno *et al.* (2001) and Dolezal *et al.* (2004).

21

22 Here, we propose to use Markov switching linear mixed models (Chaubert-Pereira *et al.*, 2010)
23 to analyze growth components of walnut saplings. In a Markov switching linear mixed model,
24 the underlying Markov chain represents both the succession and lengths of growth phases,
25 while the linear mixed models attached to each state of the Markov chain represent both the

1 effect of time-varying climatic explanatory variables and inter-individual heterogeneity. The
2 effect of climatic explanatory variables is modeled as a fixed effect and inter-individual
3 heterogeneity as a random effect. Thus, the introduction of random effects makes it possible to
4 decompose the total variability into two parts: variability due to inter-individual heterogeneity
5 and residual variability. In this study, Markov switching models were identified using semi-
6 Markov switching models, a more general family of models that encompasses as particular
7 cases Markov switching models. In a semi-Markov switching model, the length of each growth
8 phase is explicitly modeled by a dedicated parametric discrete distribution (Chaubert-Pereira *et*
9 *al.*, 2009) while in a Markov switching model, the length of each growth phase is implicitly
10 modeled by a geometric distribution, which is the unique “memoryless” discrete distribution.
11 This led us to contrast the development of saplings in understory with the development of trees
12 in managed forest stands studied by Chaubert-Pereira *et al.* (2009).

13
14 Markov switching linear mixed models were applied to the analysis of successive annual shoots
15 along walnut main stems using two possible response variables, namely (i) the number of
16 leaves per annual shoot corresponding to organogenesis alone and (ii) annual shoot length,
17 corresponding to organogenesis and elongation, and two climatic explanatory variables, namely
18 (i) average daily maximum temperature and (ii) cumulative rainfall. The building of two
19 models, one corresponding to organogenesis alone and the other corresponding to
20 organogenesis and elongation can be viewed as a way to refine the decomposition of tree
21 growth that is intrinsic to Markov switching linear mixed models. Light environments were also
22 compared for four tree categories deduced from the statistical modeling. The objective was to
23 study the relationship between the architecture of these juvenile trees growing in a mixed
24 Mediterranean forest and the two selected climatic factors and the local light environment.

25

1 The aim of this study was to assess the effects of the environment on understory saplings
2 development. In this study, the ontogeny was viewed as “a developmentally programmed
3 growth trajectory preadjusted to the most likely environments; the surrounding environments
4 mainly affect the rate at which saplings move along this trajectory.” (Yagi, 2009). Two main
5 hypotheses were tested: (i) unfavorable light environment delays tree ontogeny but without
6 impacting growth potential and, (ii) The effects of climatic factors depend on the ontogenetic
7 stage.

8

9 **Materials and methods**

10

11 **Materials**

12 **Study site**

13 The studied stand was a Mediterranean mixed forest located near Montpellier (43°40'33" N;
14 3°51'53" E) in the south of France. Mean stand elevation is 57 m and the climate is typically
15 Mediterranean (dry summer, rainfall mainly in autumn and spring) with mean annual
16 cumulative rainfall of 732 mm, an annual mean daily maximum temperature of 20.2°C and an
17 annual mean daily minimum temperature of 9.4°C for the period 1977-2006. The forest canopy
18 mainly consisted of crowns of *Quercus ilex* L., *Quercus pubescens* Willd., *Pinus halepensis*
19 Mill. and *Tilia platyphyllos* Scop.. The understory consisted of saplings of *Q. ilex*, *Q. pubescens*,
20 *T. platyphyllos*, *Celtis australis* L., *Prunus avium* L., and *Juglans regia* L.. *Hedera helix* L. and
21 *Ruscus aculeatus* L. covered the ground. The transmitted light, i.e. the percentage of incident
22 radiations not intercepted by the canopy, ranged from 5 to 35 % in the studied understory.

23

24 **Tree data set**

25 In this study, 138 saplings of walnut (*Juglans regia* L.) were measured in a natural population

1 (see Figs. 1 and 2). Since only one mother tree was present at the edge of the studied stand, we
2 suspect that most of the saplings were half-brothers. The successive annual shoots along the
3 main stems were entirely preformed, i.e. all the organs of the next-year elongated shoot were
4 present at an embryonic stage in the winter bud. The limit between two successive annual
5 shoots is marked by a zone of short internodes and scale leaves (Sabatier and Barthélémy,
6 2001) which facilitates the retrospective measurement of successive annual shoots along the
7 main stem. In March 2007 before bud break, the main stems of the trees were described by
8 annual shoot and three quantitative variables were recorded for each annual shoot: number of
9 leaves (between 2 to 12), length in cm (between 1 and 49 cm; see Figs. 1 and 2) and number of
10 branches. Of the 138 trees measured, 22 were branched. The measured trees were between 3
11 and 31 years old. All these trees were in the juvenile phase (i.e. before the first flowering
12 occurrence) the last year considered.

13
14 The very first annual shoot corresponding to the germination year was not included in the
15 analysis as it is far longer than the subsequent annual shoots and its length (range: 2-53 cm;
16 mean: 20.22 cm; standard deviation: 6.54 cm) is greatly dependent on seed mass. The
17 correlation coefficients between the length of this very first annual shoot and the length of each
18 of the five subsequent annual shoot were not significantly different from 0 (as well as between
19 the length of this very first annual shoot and the cumulative length of the - from two to five -
20 subsequent annual shoots). Hence, the length of this first annual shoot was not related to
21 subsequent tree development. This can be interpreted as the fact that the local environment had
22 a far stronger effect on the sapling development after the first year of growth than the seed
23 mass. After the germination year, the length of successive annual shoots increased along the
24 main stem. Individuals showed synchronous inter-annual fluctuations the amplitude of which
25 was roughly proportional to the growth level (i.e. average shoot length within a phase) (Fig. 1).

1 These inter-annual fluctuations were thus assumed to be mainly of climatic origin.

2 **Meteorological data set**

2 Data for daily maximum temperature (°C) and cumulative rainfall (in mm) were provided by
3 INRA-Montpellier (Lavalette site, a nearby meteorological field station) for the period from
4 1977 to 2006 (the last year considered).

5

6 **Measuring light and leaf mass area**

7 Light availability in the understory was quantified by hemispherical photography. Photographs
8 were taken in May 2007 after leaf expansion of canopy species, above the terminal bud of the
9 main stem of each studied tree. The photographs were taken using a horizontally-leveled digital
10 camera (CoolPix 8400, Nikon) mounted on a tripod and aimed at the zenith. A fish-eye lens
11 providing a 180° field of view was employed (Coolpix FC-E9, Nikon). These photographs were
12 analyzed for canopy openness (i.e. percentage of open sky) and percentage of light transmitted
13 by the forest canopy during the growing season (i.e. April-October) using GLA2 software
14 (Frazer *et al.*, 1999). These two indicators are commonly used to compare sapling light
15 environments (Kobe and Hogarth, 2007).

16

17 Leaf dry mass per unit leaf area (LMA) was computed from a 5 cm² piece of lamina (i.e.
18 without the leaflet midrib) taken from one leaflet per tree. These leaflets were sampled in
19 similar positions within the 2007 main stem shoot. This sampling was completed in a single
20 two-hour period during the 2007 summer. LMA computed in this manner is generally
21 considered as a reliable indicator of leaf structure (Niinemets 1999).

22

23 **Methods**

1 **Markov and semi-Markov switching linear mixed models**

2 In the analysis we considered two possible tree response variables: (i) the number of leaves per
3 annual shoot and (ii) the length of the annual shoot in cm. In cases where climatic explanatory
4 variables are available, a statistical model for analyzing this type of tree growth data should be
5 able to model jointly:

- 6 - the succession of roughly stationary growth phases that are asynchronous between
7 individuals,
- 8 - the effect of time-varying climatic explanatory variables,
- 9 - inter-individual heterogeneity.

10 We therefore chose to build Markov switching linear mixed models. This family of statistical
11 models broadens the family of Markov switching linear models; see Frühwirth-Schnatter
12 (2006) for an overview of Markov switching models. A Markov switching linear mixed model
13 combines:

- 14 - a non-observable J -state Markov chain which represents the succession of growth
15 phases and their lengths where each state of the Markov chain represents a growth phase.

16 A J -state Markov chain is defined by two subsets of parameters:

- 17 1. Initial probabilities $(\pi_j; j = 1, \dots, J)$ to model which is the first phase occurring
18 in the sequence measured for a tree,
- 19 2. Transition probabilities $(p_{ij}; i, j = 1, \dots, J)$ to model the succession of growth
20 phases and their lengths along the main stem.

- 21
- 22 - J linear mixed models, each one attached to a state of the underlying Markov chain.
23 Each linear mixed model represents, in the corresponding growth phase, both the effect
24 of time-varying climatic explanatory variables as fixed effects and inter-individual

1 heterogeneity as a random effect; see Verbeke and Molenberghs (2000) for an overview
2 of linear mixed models applied to longitudinal data.

3
4 The observed annual shoot length $Y_{a,t}$ for tree a being in state j at year t is modeled by the
5 following linear mixed model:

$$6 \quad Y_{a,t} = \beta_{j1} + \beta_{j2}X_{2,t} + \beta_{j3}X_{3,t} + \tau_j\xi_{a,j} + \varepsilon_{a,t},$$

7 where:

8 - $\beta_{j1} + \beta_{j2}X_{2,t} + \beta_{j3}X_{3,t}$ is the contribution of the fixed effects for state j , β_{j1} the intercept
9 which represents the average growth level, β_{j2} the regression parameter for average daily
10 maximum temperature, $X_{2,t}$ the average daily maximum temperature common to all the trees
11 that is relevant for year t , β_{j3} the regression parameter for cumulative rainfall, $X_{3,t}$ the
12 cumulative rainfall common to all the trees that is relevant for year t . It should be noticed that
13 the tree response to time-varying climatic factors is most often delayed and “smoothed”.

14 - $\xi_{a,j} \sim N(0,1)$ is the random effect attached to tree a being in state j and τ_j the standard
15 deviation induced by the inter-individual heterogeneity in state j . The random effects are
16 assumed to follow the standard Gaussian distribution $N(0,1)$. Individual tree status is assumed
17 to be different in each growth phase. For a given individual, the average shoot length within a
18 phase can be higher than that of the “average tree” then lower in the following phase.

19 - $\varepsilon_{a,t} | S_{a,t} = j \sim N(0, \sigma_j^2)$ is the error term corresponding to tree a being in state j at year t
20 and σ_j^2 is the residual variance.

21
22 When considering the number of leaves per annual shoot, the linear mixed model is similar
23 except for the cumulative rainfall fixed effect which is not incorporated; see the Results for

1 justifications. The introduction of random effects makes it possible to decompose the total
2 variability Γ_j^2 in state j into two parts:

$$3 \quad \Gamma_j^2 = \tau_j^2 + \sigma_j^2,$$

4 where τ_j^2 is the variability due to inter-individual heterogeneity and σ_j^2 is the residual
5 variability. The proportion of inter-individual heterogeneity is defined by the ratio between the
6 random effect variance τ_j^2 and the total variance Γ_j^2 .

7
8 Markov switching linear mixed models were estimated using a Monte Carlo expectation
9 maximization (MCEM)-like algorithm (Chaubert-Pereira *et al.*, 2010). At convergence of the
10 iterative estimation algorithm, the median predicted random effects were computed for each
11 individual based on the random effects predicted for each state in each sampled state sequence.
12 The most probable state sequence given the median predicted random effects was computed for
13 each observed sequence using a Viterbi-like algorithm. This restored state sequence can be
14 considered as the optimal segmentation of the corresponding observed sequence into sub-
15 sequences, each corresponding to a given growth phase.

16
17 Standard errors for the regression parameters of Markov switching linear mixed models were
18 computed using a Monte Carlo version of the Louis method (Chaubert-Pereira *et al.*, 2010). In
19 our case of non-ergodic models (a model is said to be ergodic if each state is visited many
20 times), we chose also to compute $\beta_{jk} \times \text{mad}_j(X_k)$ for each state j where $\text{mad}_j(X_k)$ is the
21 mean absolute deviation of the climatic explanatory variable k in state j . This indicator gives
22 empirical evidence of the significant or non-significant character of the corresponding climatic
23 effect.

24

1 A useful generalization of Markov switching linear mixed models lies in the class of semi-
2 Markov switching linear mixed models (Chaubert-Pereira *et al.*, 2010) where the underlying
3 Markov chain is replaced by an underlying semi-Markov chain. In a semi-Markov chain, the
4 process moves out of a given state according to a Markov chain with self-transition probability
5 in nonabsorbing states $\tilde{p}_{ii} = 0$. This embedded Markov chain represents transitions between
6 distinct states except in the absorbing state case. A state is said to be absorbing if, after entering
7 this state, it is impossible to leave it. An explicit occupancy (or sojourn time) distribution is
8 attached to each nonabsorbing state to model the growth phase length in number of years. The
9 whole (embedded Markov chain + explicit state occupancy distributions) constitutes a semi-
10 Markov chain. The mechanism associated with a semi-Markov chain can be described as
11 follows: suppose that between two consecutive times, a transition occurred between state i and
12 state j with probability \tilde{p}_{ij} . The process remains in state j for a period u determined randomly
13 by the corresponding state occupancy distribution $\{d_j(u); u = 1, 2, \dots\}$. Then the process moves to
14 another state according to the transition distribution attached to state j ($\tilde{p}_{jk}; k = 1, \dots, J$).
15 Markov and semi-Markov switching linear mixed models are implemented in a module that
16 will be integrated in the OpenAlea software platform for plant modeling (Pradal *et al.*, 2008).
17 Complementary statistical analyses were performed using R. Markov and semi-Markov
18 switching linear mixed models are formally defined in the Appendix.

19

20 **Results**

21 **Choice of climatic explanatory variables**

22 In walnut, organogenesis occurs from early April to late May of the previous year and organ
23 elongation occurs the current year during the same period (Sabatier *et al.*, 2003b). We chose to
24 build two statistical models, the first corresponding to organogenesis alone where the tree

1 response variable is the number of leaves per annual shoot, and the second corresponding to
2 organogenesis and elongation where the tree response variable is the annual shoot length. Since
3 both the number of leaves and the length of successive annual shoots increased along the main
4 stem (Fig. 1), these variables could not be considered as stationary and non-ergodic models, in
5 our case “left-right” models (see below), were built. We chose to use average daily maximum
6 temperature and cumulative rainfall and to center these climatic explanatory variables; see
7 Fitzmaurice *et al.* (2004) for discussion of the centering issue. Because of the centering, the
8 intercept β_{j1} is directly interpretable as the average number of leaves per annual shoot in
9 growth phase j . In a first step, we tested different climatic explanatory variables for the
10 “number of leaves” model. We did not identify a significant cumulative rainfall effect for this
11 model whatever the period tested. The maximum temperature explanatory variable selected for
12 the “number of leaves” model was then incorporated in the “annual shoot length” model and
13 other climatic explanatory variables were tested in a second step for this model.

14
15 In order to determine the most significant period for average daily maximum temperature, a
16 “left-right” three-state Markov switching linear model (i.e. without random effects for
17 modeling inter-individual heterogeneity) composed of two transient states followed by a final
18 absorbing state was first estimated for different periods covering the organogenesis period on
19 the basis of the number of leaves per annual shoot; see below for the choice of the number of
20 states of the underlying Markov chain. A state is said to be transient if after leaving this state it
21 is impossible to return to it. In a “left-right” model, the states are therefore ordered and each
22 state can be visited at most once. In this sensitivity analysis, we selected the period on the basis
23 of observed data log-likelihood (results not shown). Centered average daily maximum
24 temperature in °C from April to May of the previous year, which covered one organogenesis
25 period, was finally selected.

1
2 A “left-right” three-state Markov switching linear model was then estimated on the basis of
3 annual shoot length of walnuts for different temperature periods preceding the end of
4 elongation and for different rainfall periods covering organogenesis and/or elongation. In
5 addition to centered average daily maximum temperature in °C from April to May of the
6 previous year, centered cumulative rainfall (in mm) from June to December of the previous
7 year, which covered the period during which carbohydrate reserves were accumulated for bud
8 break (Lacointe *et al.*, 1995), was finally selected.

9

10 **Population properties**

11 We first estimated a “left-right” three-state semi-Markov switching linear model (i.e. without
12 random effects). Since the estimated state occupancy distributions for states 1 and 2 were quite
13 close to 1-shifted geometric distributions (see the shape parameter r and the mean and standard
14 deviation of the estimated state occupancy distributions in Table 1), we chose to estimate
15 Markov switching linear models based on simple Markov chains. For both the “number of
16 leaves” and the “annual shoot length” models, the log-likelihood of the observed sequences for
17 the Markov switching linear model was close to the log-likelihood of the observed sequences
18 for the semi-Markov switching linear model (Table 1). We thus chose to use Markov switching
19 models for subsequent analyses. The identification of approximately geometrically distributed
20 phase length can be interpreted as the consequence of an opportunistic growth process mainly
21 driven by local environment constraints. It should be recalled that the geometric distribution is
22 the unique discrete “memoryless” distribution which means that the fact that a tree was in a
23 given growth phase for a certain number of years in no way affects the probability that it will
24 stay for k further years in this phase or make a transition to the subsequent growth phase.

1 “Left-right” three-state Markov switching linear mixed models were estimated on the basis of
2 either the number of leaves per annual shoot or the annual shoot lengths of walnuts. We applied
3 the practical approach discussed in Guédon *et al.* (2007) to determine the number of states of
4 the underlying Markov chain. For each tree response variable, the iterative estimation algorithm
5 was initialized with a “left-right” model such that $\pi_j > 0$ for each state j , $p_{ij} = 0$ for $j < i$ and
6 $p_{ij} > 0$ for $j \geq i$. The fact that states 1 and 2 are the only possible initial states (with $\pi_1 = 0.79$
7 and $\pi_2 = 0.21$ at convergence for the “number of leaves” model and $\pi_1 = 0.68$ and $\pi_2 = 0.32$
8 for the “annual shoot length” model) and that state 2 cannot be skipped (i.e. $p_{13} = 0$ at
9 convergence) is the result of the iterative estimation procedure; see Fig. 3 for the “number of
10 leaves” model and Fig. 4 for the “annual shoot length” model. This deterministic succession of
11 states supports the assumption of a succession of growth phases. For the two estimated models,
12 we checked that sub-sequences extracted for the three states according to the optimal
13 segmentation do not exhibit residual trends within the corresponding phase; see Guédon *et al.*
14 (2007) for details concerning methods used to assess the stationarity assumption.

15
16 Whilst the transition probabilities are different for the two estimated Markov chains ($p_{11} = 0.81$
17 and $p_{22} = 0.94$ for the “number of leaves” model while $p_{11} = 0.91$ and $p_{22} = 0.9$ for the
18 “annual shoot length” model), mean times up to the first occurrence of state 3 computed from
19 model parameters were fairly similar ($\eta_3 = 19.66$ for the “number of leaves” model and
20 $\eta_3 = 16.9$ for the “annual shoot length” model); see Table 2. This may be explained by the fact
21 that the first two growth phases are less separated than the last two growth phases, particularly
22 for the “number of leaves” model; see below the analysis of the marginal observation
23 distributions. In this respect, the fact that a certain proportion of the trees started directly in

1 state 2 (with $\pi_2 = 0.21$ for the “number of leaves” model and $\pi_2 = 0.32$ for the “annual shoot
2 length” model) should not be given too much meaning.

3
4 It should be noted that the succession of growth phases is characterized by an increase in the
5 number of leaves per annual shoot or in the length of the annual shoot and by an increase in the
6 variability of annual shoot length; see Table 3 and Fig. 3a for the “number of leaves” model and
7 Table 4 and Fig. 4a for the “annual shoot length” model. The marginal observation distribution
8 of the linear mixed model attached to growth phase j is the Gaussian distribution $N(\mu_j, \Gamma_j^2)$
9 with $\mu_j = \beta_{j1} + \beta_{j2}E_j(X_2)$ for the “number of leaves” model and
10 $\mu_j = \beta_{j1} + \beta_{j2}E_j(X_2) + \beta_{j3}E_j(X_3)$ for the “annual shoot length” model, where $E_j(X_2)$ is the
11 mean centered average daily maximum temperature in growth phase j and $E_j(X_3)$ is the mean
12 centered cumulative rainfall in growth phase j . Depending on the tree response variable of
13 interest, the marginal observation distribution represents either the number of leaves per annual
14 shoot or the annual shoot length in growth phase j . The marginal observation distributions for
15 states 2 and 3 are more separated (less overlap between the marginal observation distributions)
16 than the marginal observation distributions for states 1 and 2, particularly for the “annual shoot
17 length” model; compare the mean difference $\mu_{j+1} - \mu_j$ with the standard deviations Γ_j and
18 Γ_{j+1} for the two estimated Markov switching linear mixed models in Tables 3 and 4 and see
19 also Figs. 3a and 4a. The two main tree categories thus consisted of trees not reaching growth
20 phase 3 and trees reaching growth phase 3 (on the basis of the “annual shoot length” model). In
21 particular, all branched trees belonged to this latter category.

22
23 For the “number of leaves” model, the temperature effect was not considered as significant; see
24 Table 3. For the “annual shoot length” model, the temperature effect and the cumulative rainfall

1 effect were weak in the first growth phase (of slowest growth) while they were stronger in the
2 last two growth phases (less in the second than in the third phase); see Table 4. This behavior is
3 similar to that observed in Corsican pine and sessile oak when analyzing the effect of
4 cumulative rainfall alone on annual shoot length using a semi-Markov switching linear mixed
5 model (Chaubert-Pereira *et al.*, 2009). However, the temperature effect was more marked than
6 the cumulative rainfall effect in each growth phase. This difference increased markedly in
7 growth phase 3. The ratio between the temperature effect and the cumulative rainfall effect was
8 approximately 1.9, 3.1 and 3.6 in growth phases 1, 2 and 3, respectively; see the ratio
9 $\{\beta_{j_2} \times \text{mad}_j(X_2)\} / \{\beta_{j_3} \times \text{mad}_j(X_3)\}$ in Table 4. The two selected climatic factors thus had an
10 impact on elongation but had little effect on organogenesis.

11

12 **Individual behavior**

13 The proportion of inter-individual heterogeneity was greater in the first two growth phases with
14 more than 47% for the “number of leaves” model and more than 43% for the “annual shoot
15 length” model, before decreasing markedly in the last growth phase (approx. 29% and 26%,
16 respectively); see Tables 3 and 4. Because of the genetic proximity between the trees, we
17 suspect that a large part of the inter-individual heterogeneity may be explained by the
18 heterogeneous environment of the trees. This local environment was a stronger constraint in the
19 first two growth phases than in the last where the comparatively larger trees were less sensitive
20 to competitions for light and mineral nutrients. The growth phases identified by the estimated
21 Markov switching linear mixed models are thus not only defined by the average number of
22 leaves per annual shoot or annual shoot length but also by the amplitude of synchronous
23 fluctuations between individuals due to climatic factors and by the proportion of inter-
24 individual heterogeneity.

25

1 For the “number of leaves” and “annual shoot length” models, all the 22 branched walnuts,
2 except one for the “number of leaves” model, reached growth phase 3 whereas only about 25%
3 of the 116 unbranched walnuts reached this growth phase (25 of 116 for the “number of leaves”
4 model and 30 of 116 for the “annual shoot length” model); see Table 5. For the branched trees,
5 the time interval between the beginning of growth phase 3 and the first branching occurrence
6 was extracted on the basis of the optimal segmentation of the observed sequence computed
7 using the “annual shoot length” model. The mean time interval was 3.33 years (standard
8 deviation: 2.85 years) indicating that the first branching occurrence requires a time interval with
9 respect to the beginning of the third growth phase.

10

11 **Light environment effect**

12 The growth phase in 2006 (the last year considered), according to the optimal segmentation of
13 the observed sequence computed using the “annual shoot length” model, and the
14 absence/presence of branches, was used to define four tree categories: (i) trees in growth phase
15 1, (ii) trees in growth phase 2, (iii) unbranched trees in growth phase 3 and (iv) branched trees
16 in growth phase 3. Canopy openness, transmitted light and LMA (during the 2007 growth
17 period) were compared between these four tree categories (Fig. 5). According to the Tukey test,
18 the values computed for the trees in growth phases 1 and 2 were not significantly different for
19 each light environment indicator. These trees had lower values for canopy openness,
20 transmitted light and LMA than trees in growth phase 3. Trees in the first two growth phases
21 were therefore growing in a more shaded environment than trees in growth phase 3. Canopy
22 openness and transmitted light values computed for the unbranched and branched trees in
23 growth phase 3 were not significantly different. LMA was the only indicator with significantly
24 different values between unbranched and branched trees in growth phase 3.

25

1 The average tree profile (i.e. $\beta_{s_{a,t}1} + \beta_{s_{a,t}2} \times$ centered average daily maximum temperature +
 2 $\beta_{s_{a,t}3} \times$ centered cumulative rainfall for each year t) and the predicted tree profile (i.e. average
 3 tree profile value + $\tau_{s_{a,t}} \xi_{a,s_{a,t}}$ for each year t) were computed for each tree a on the basis of the
 4 optimal segmentation of the observed sequence and the predicted random effects computed
 5 using the “annual shoot length” model. The regression parameters β_{j1} , β_{j2} and β_{j3} describe
 6 patterns of change in the mean response over years (and their relationship to explanatory
 7 variables) in the walnut population, while $\tau_j \xi_{a,j}$ describes the deviation in growth phase j of
 8 the a th tree profile with reference to the average tree profile; see Fig. 6 for the predicted tree
 9 profiles of 6 representative individuals. Four types of tree growth profile were identified:

- 10 ▪ Trees characterized by regularly spaced phase changes; see for instance tree 123 where
 11 the lengths of growth phases 1 and 2 were similar (4 years in growth phase 1 and 5
 12 years in growth phase 2);
- 13 ▪ Trees with constant slow growth; see tree 86 which remained in growth phase 1. These
 14 trees remained in unfavorable local conditions (dark understory) throughout the period
 15 considered;
- 16 ▪ Trees with a long period of slow growth followed by a rapid increase in the annual
 17 shoot length corresponding to growth phase changes; see trees 58 and 133. We suspect
 18 that these phase changes were due to a rapid modification of the understory light
 19 environment enabling a rapid growth increase followed by first branching (see tree 133)
 20 despite a long period of slow growth. In this type of tree growth profile, the predicted
 21 random effects in growth phases 1 and 3 did not seem to be related and the long phase
 22 of slow growth did not seem to affect subsequent tree development;
- 23 ▪ Trees with a short period of slow growth before a long period of strong growth; see
 24 trees 130 and 124. These trees tended to branch.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Tree status with reference to the average tree may be common to all the growth phases but modulated in terms of deviations between growth phases; see for example trees 130 and 124 in Fig. 6. Tree status may also be different in each growth phase; see for example trees 58 and 133 in Fig. 6. The more general assumption of a random effect attached to each growth phase is thus more representative of walnut behavior than the assumption of a random effect common to all growth phases.

Discussion

Walnuts in understory: a shade avoidance response pattern

Trees in the third growth phase in 2006 (the last year considered) were growing in a more luminous environment and had greater LMA than trees in the first two growth phases. Light availability is certainly the main factor that explains a tree's ability to reach the third phase of strongest growth; see Niinemets (1999) for the link between LMA and light availability. Walnuts showed opportunistic growth in understory characterized by their capability to increase their apical growth suddenly in response to a change in the local light environment, for instance arising from a treefall gap. This result is supported by a cross-sectional study of a subsample of the same walnut data set (Taugourdeau and Sabatier, 2010). This behavior corresponds to a shade avoidance strategy (Henry and Aarssen, 2001). Because of the sample size (138 trees) and the marked heterogeneity of the studied mixed forest, the local environment of the trees and its time variations can be considered as sufficiently randomized. This explains the emergence of geometric phase length distributions for both the “number of leaves” and the “annual shoot length” models. This contrasts with the successive growth phases with bell-shaped length distributions such as in Corsican pines and sessile oaks in managed forest stands (Chaubert-Pereira *et al.*, 2009). In this latter case, the well-defined successive growth phases

1 were mainly the expression of endogenous equilibriums while in the case studied here, the
2 phase length was the result of an opportunistic process corresponding to a tree's plastic response
3 to environmental changes.

4
5 The transition to the third growth phase was related to a reduction in inter-individual
6 heterogeneity. Because of the genetic proximity between the trees (same mother), this may be
7 interpreted as the effect of the local environment on the growth of each tree. In the first two
8 growth phases, young trees, whatever their local environment, were mixed up with older trees
9 in unfavorable local environments while only trees in reasonably favorable local environments
10 reached the third growth phase. This is somewhat similar to the environmental filter concept:
11 see Diaz *et al.* (1999). A complementary explanation lies in the fact that large trees (in the third
12 growth phase) were less sensitive than smaller trees to the local environment.

13
14 Markov switching linear mixed model allowed us to highlight the fact that a long period of
15 suppression has no negative effect on subsequent growth after canopy opening. This is
16 illustrated for instance by tree 58 (Fig. 6) which grew faster than the average tree in the third
17 growth phase despite a long period of suppression corresponding to the first growth phase.
18 Wright *et al.* (2000) came to the same conclusion on the basis of tree ring data.

19
20 Walnut exhibits strong apical dominance in shaded environments since lateral buds only
21 develop after canopy opening (Fig. 7a). This corresponds to a shade avoidance strategy
22 characterized by a major investment in vertical growth to reach higher irradiance strata and a
23 marked response to changes in the light environment (Wright *et al.*, 2000; Henry and Aarssen
24 2001). This is consistent with the “heliophil status” of walnut. Henry and Aarssen (2001)
25 suggest that a shade tolerance strategy is specific of species growing in the deepest shaded

1 environments and a shade avoidance strategy is specific of species growing in environments
2 characterized by high vertical and horizontal heterogeneity of light availability such as in the
3 studied understory. Light quality was not quantified in this study but is known to be related to
4 the shade avoidance strategy (Franklin and Whitelam, 2005).

5
6 Some shade-tolerant species accommodate to a shaded environment by an important lateral
7 development. For instance, fir has long lifespan branches and needles (Mori *et al.*, 2008) in a
8 shaded environment. The response of walnut to shade is very different since it consists of
9 producing juvenile deciduous compound leaves characterized by dentate leaflets.

10

11 **Effects of temperature and rainfall on annual growth**

12 The effect of average daily maximum temperature and cumulative rainfall was weak in the first
13 phase where shade drastically limited tree development (Fig. 7b). It then increased markedly in
14 the subsequent growth phases. This is consistent both with the results of Mediavilla and
15 Escudero (2004) concerning the response to drought of trees at different stages of development
16 and with the facilitation hypothesis proposed by Holmgren (2000); see also Sack (2004) for a
17 discussion about ecological consequences.

18

19 The proposed modeling approach highlighted the positive effect of average daily maximum
20 temperature during shoot organogenesis on the next-year assimilative leaf number (one-year-
21 delayed response). The relationship between temperature and organogenesis has been well
22 established and is the basis of the thermal time concept applied mainly to annual plants
23 (Bonhomme, 2000). The effect of temperature has generally been studied in plants where shoot
24 elongation occurs just after shoot organogenesis, making it difficult to separate the effect on
25 organogenesis from the effect on elongation. In the current study of a temperate perennial

1 species, the one year delay between shoot organogenesis and elongation enabled to clearly
2 decipher the temperature effect. The delay between the change in environment and the plant's
3 response may be an adaptative drawback since such a delayed response may not be appropriate
4 in case of rapid change of the environment (Valladares *et al.*, 2007).

5
6 The temperature during organogenesis influences the next-year assimilative leaf number but
7 also, and to a greater extent, the annual shoot length. This result may at first sight appear to be
8 counterintuitive since we did not identify other periods where annual shoot length was affected
9 by average daily maximum temperature. We therefore suspect that the average daily maximum
10 temperature during organogenesis not only influences the number of organs but also the
11 number of cells, and consequently the length of the corresponding internodes; see Ripetti *et al.*
12 (2008) and references therein.

13
14 **Modeling forest structure and successional dynamics in a context of climate change**

15 Markov switching linear mixed models were applied to identify and characterize the effects of
16 ontogeny, climatic factors and local environment on tree establishment. At a coarser scale, the
17 outputs of these individual models could be incorporated into models of forest structure and
18 successional dynamics:

- 19 ▪ Predictions about possible rank reversal (i.e. changes in interspecific competition) in
20 mixed forest due to climate change can also be made; see Sánchez-Gómez *et al.* (2008)
21 for a discussion about the ecological consequences of rank reversal. For example, a
22 warmer and drier Mediterranean climate (Christensen *et al.*, 2007) would have
23 antagonist effects on walnut growth: the increased temperature would have a positive
24 effect but the decreased rainfall would have a negative effect. Local rainfall and
25 temperature predictions for 2050 (ARPEGE/IFS model; Déqué *et al.*, 1994) would have

1 a positive effect on the studied walnuts in the third growth phase (+1 leaf and +9 cm per
2 annual shoot) because of the relative importance of temperature on growth. These
3 predictions assume that the selected climatic factors are not subject to threshold effects.

4 ■ The structure of a population can be characterized at a given date by the proportion of
5 trees in the different growth phases. We expect this classification based on a
6 decomposition of tree growth components to be more robust than alternative
7 classifications based on direct measurements (e.g. diameter measured at breast height).

8

9 **Acknowledgments**

10 The authors are grateful to Michaël Gueroult for his help in collecting field data, to INRA for
11 freely providing meteorological data, to Yves Caraglio, Pierre Couteron and Nick Rowe for
12 useful comments, to the handling Editor and an anonymous reviewer for their helpful
13 comments that led to an improvement in the presentation of this paper.

14

Literature cited

- Barthélémy D, Caraglio Y. 2007.** Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. *Annals of Botany* **99**: 375-407.
- Bonhomme R. 2000.** Bases and limits to using 'degree.day' units. *European Journal of Agronomy* **13**: 1-10.
- Chaubert-Pereira F, Caraglio Y, Lavergne C, Guédon Y. 2009.** Identifying ontogenetic, environmental and individual components of forest tree growth. *Annals of Botany* **104**: 883-896.
- Chaubert-Pereira F, Guédon Y, Lavergne C, Trottier C. 2010.** Markov and semi-Markov switching linear mixed models used to identify forest tree growth components. *Biometrics* **66**(3): 753-762.
- Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli RK, Kwon W-T, Laprise R, Magaña Rueda V, Mearns L, Menéndez CG, Räisänen J, Rinke A, Sarr A, Whetton P. 2007.** Regional Climate Projections. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL. (eds.)]. Cambridge University Press, Cambridge.
- Déqué M, Dreveton C, Braun A, Cariolle D. 1994.** The ARPEGE/IFS atmosphere model: a contribution to the French community climate modelling. *Climate Dynamics* **10**: 249-266.
- Diaz S, Cabido M, Zak M, Carretero EM, Aranibar J. 1999.** Plant functional traits, ecosystem structure and land-use history along a climatic gradient in central-western Argentina. *Journal of Vegetation Science* **10**: 651-660.
- Dolezal J, Ishii H, Vetrova VP, Sumida A, Hara T. 2004.** Tree growth and competition in a *Betula platyphylla-Larix cajanderi* post-fire forest in central Kamchatka. *Annals of Botany* **94**: 333-343.
- Fitzmaurice GM, Laird NM, Ware JH. 2004.** *Applied longitudinal analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ: John Wiley & Sons.
- Franklin KA, Whitelam GC. 2005.** Phytochromes and shade-avoidance responses in plants. *Annals of Botany* **96**: 169 -175.
- Frazer GW, Canham CD, Lertzman KP. 1999.** Gap Light Analyzer (GLA), Version 2.0: Imaging software to extract canopy structure and gap light transmission indices from true-colour fisheye photographs, users manual and program documentation. Simon Fraser University, Burnaby BC, and the Institute of Ecosystem Studies, Millbrook, NY.
- Frühwirth-Schnatter S. 2006.** *Finite Mixture and Markov Switching models*. Springer Series in Statistics. New York: Springer.
- Guédon Y, Caraglio Y, Heuret P, Lebarbier E, Meredieu C. 2007.** Analyzing growth components in trees. *Journal of Theoretical Biology* **248**(3): 418-447.

- Henry HAL, Aarssen LW. 2001.** Inter- and intraspecific relationships between shade tolerance and shade avoidance in temperate trees. *Oikos* **93**: 477-487.
- Holmgren M. 2000.** Combined effects of shade and drought on tulip poplar seedlings: trade-off in tolerance or facilitation? *Oikos* **90**: 67-78.
- Kobe RK, Hogarth LJ. 2007.** Evaluation of irradiance metrics with respect to predicting sapling growth. *Canadian Journal of Forest Research* **37**: 1203-1213.
- Kulkarni VG. 1995.** *Modeling and Analysis of Stochastic Systems*. London: Chapman & Hall.
- Lacointe A, Kajji A, Daudet FA, Archer P, Frossard JS. 1995.** Seasonal-Variation of Photosynthetic Carbon Flow-Rate into Young Walnut and Its Partitioning among the Plant Organs and Functions. *Journal of Plant Physiology* **146**: 222-230.
- Mediavilla S, Escudero A. 2004.** Stomatal responses to drought of mature trees and seedlings of two co-occurring Mediterranean oaks. *Forest Ecology and Management* **187**: 281-294.
- Miner B, Sultan S, Morgan S, Padilla D, Relyea R. 2005.** Ecological consequences of phenotypic plasticity. *Trends in Ecology & Evolution* **20**: 685-692.
- Mori AS, Mizumachi E, Sprugel DG. 2008.** Morphological acclimation to understory environments in *Abies amabilis*, a shade- and snow-tolerant conifer species of the Cascade Mountains, Washington, USA. *Tree Physiology* **28**: 815-824.
- Niinemets U. 1999.** Components of leaf dry mass per area - thickness and density - alter leaf photosynthetic capacity in reverse directions in woody plants. *New Phytologist* **144**: 35-47.
- Pinno BD, Lieffers VJ, Stadt KJ. 2001.** Measuring and modelling the crown and light transmission characteristics of juvenile aspen. *Canadian Journal of Forest Research* **31**: 1930-1939.
- Pradal C, Dufour-Kowalski S, Boudon F, Fournier C, Godin C. 2008.** OpenAlea: A visual programming and component-based software platform for plant modeling. *Functional Plant Biology* **35**(9-10): 751-760.
- Ripetti V, Escoute J, Verdeil JL, Costes E. 2008.** Shaping the shoot: the relative contribution of cell number and cell shape to variations in internode length between parent and hybrid apple trees. *Journal of Experimental Botany* **59**: 1399-1407.
- Sabatier S, Baradat P, Barthélémy D. 2003a.** Intra- and interspecific variations of polycyclism in young trees of *Cedrus atlantica* (Endl.) Manetti ex. Carriere and *Cedrus libani* A. Rich (Pinaceae). *Annals of Forest Science* **60**: 19-29.
- Sabatier S, Barthélémy D. 2001.** Bud structure in relation to shoot morphology and position on the vegetative annual shoots of *Juglans regia* L. (Juglandaceae). *Annals of Botany* **87**(1): 117-123.
- Sabatier S, Barthélémy D, Ducouso I. 2003b.** Periods of organogenesis in mono- and bicyclic annual shoots of *Juglans regia* L. (Juglandaceae). *Annals of Botany* **92**: 231-238.

- Sack L. 2004.** Responses of temperate woody seedlings to shade and drought: do trade-offs limit potential niche differentiation? *Oikos* **107**: 110-127.
- Sánchez-Gómez D, Zavala MA, Van Schalkwijk DB, Urbieto IR, Valladares F. 2008.** Rank reversals in tree growth along tree size, competition and climatic gradients for four forest canopy dominant species in Central Spain. *Annals of Forest Science* **65**: -.
- Segura V, Cilas C, Costes E. 2008.** Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modelling of repeated spatial and temporal measures. *New Phytologist* **178**: 302-314.
- Taugourdeau O, Sabatier S. 2010.** Limited plasticity of shoot preformation in response to light by understorey saplings of common walnut (*Juglans regia*). *AoB PLANTS* **2010**: plq022, doi:10.1093/aobpla/plq022.
- Valladares F, Gianoli E, Gomez JM. 2007.** Ecological limits to plant phenotypic plasticity. *New Phytologist* **176**: 749-763.
- Verbeke G, Molenberghs G. 2000.** *Linear mixed models for longitudinal data*. Springer Series in Statistics. New York: Springer.
- Wright EF, Canham CD, Coates KD. 2000.** Effects of suppression and release on sapling growth for 11 tree species of northern, interior British Columbia. *Canadian Journal of Forest Research* **30**: 1571-1580.
- Yagi, T. 2009.** Ontogenetic strategy shift in sapling architecture of *Fagus crenata* in the dense understorey vegetation of canopy gaps created by selective cutting. *Canadian Journal of Forest Research* **39**(6): 1186-1196.

1 **Appendix**

2 Let $\{S_t\}$ be a discrete-time Markovian model with finite-state space $\{1, \dots, J\}$; see Kulkarni
3 (1995) for a general reference about Markov and semi-Markov models.

4

5 **Markov chains**

6

7 A J -state Markov chain $\{S_t\}$ is defined by the following parameters:

- 8 - initial probabilities $\pi_j = P(S_1 = j)$ with $\sum_j \pi_j = 1$;
- 9 - transition probabilities $p_{ij} = P(S_t = j \mid S_{t-1} = i)$ with $\sum_j p_{ij} = 1$.

10 The implicit occupancy (or sojourn time) distribution of a nonabsorbing state j is the “1-
11 shifted” geometric distribution with parameter $1 - p_{jj}$

$$12 \quad d_j(u) = (1 - p_{jj})p_{jj}^{u-1}, \quad u = 1, 2, \dots$$

13 This is the unique discrete memoryless distribution.

14

15 **Semi-Markov chains**

16

17 A useful generalization of Markov chains lies in the class of semi-Markov chains, in which the
18 process moves out of a given state according to an embedded Markov chain with self-transition
19 probability in nonabsorbing states $\tilde{p}_{jj} = 0$ and where the time spent in a given nonabsorbing
20 state is modeled by an explicit occupancy distribution.

21

22 A J -state semi-Markov chain $\{S_t\}$ is defined by the following parameters:

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

1 - transition probabilities

2 ▪ nonabsorbing state i : for each $j \neq i$, $\tilde{p}_{ij} = P(S_t = j | S_t \neq i, S_{t-1} = i)$ with

3 $\sum_{j \neq i} \tilde{p}_{ij} = 1$ and $\tilde{p}_{ii} = 0$ by convention,

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

5

6 An explicit occupancy distribution is attached to each nonabsorbing state:

7 $d_j(u) = P(S_{t+u+1} \neq j, S_{t+u-v} = j, v = 0, \dots, u-2 | S_{t+1} = j, S_t \neq j), \quad u = 1, 2, \dots$

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

9 $P(S_{t+1} \neq j, S_{t-v} = j, v = 0, \dots, t-1) = d_j(t)\pi_j$.

10

11 We define as possible parametric state occupancy distributions negative binomial distributions

12 with an additional shift parameter d ($d \geq 1$) which defines the minimum sojourn time in a given

13 state. The negative binomial distribution with parameters d , r and p , $NB(d, r, p)$, where r is a

14 real number ($r > 0$) and $0 < p \leq 1$, is defined by

15 $d_j(u) = \binom{u-d+r-1}{r-1} p^r q^{u-d}, \quad u = d, d+1, \dots$

16

17 A Markov chain can be reparameterized as a semi-Markov chain such that for each

18 nonabsorbing state i , $\tilde{p}_{ij} = p_{ij}/(1-p_{ii})$ for $j = 1, \dots, J, j \neq i$ and the associated explicit

19 occupancy distribution is the “1-shifted” geometric distribution $NB(1, 1, 1-p_{ii})$.

20

1 **Hidden Markov chains**

2
3 A hidden Markov chain can be viewed as a pair of stochastic processes $\{S_t, Y_t\}$ where the
4 “output” process $\{Y_t\}$ is related to the “state” process $\{S_t\}$, which is a finite-state Markov chain,
5 by a probabilistic function or mapping denoted by f (hence $Y_t = f(S_t)$). Since the mapping f is
6 such that a given output may be observed in different states, the state process $\{S_t\}$ is not
7 observable directly but only indirectly through the output process $\{Y_t\}$. This output process
8 $\{Y_t\}$ is related to the Markov chain $\{S_t\}$ by the observation (or emission) probabilities
9 $b_j(y) = P(Y_t = y | S_t = j)$ with $\int b_j(y) dy = 1$ in the case of a continuous output process. The
10 definition of observation probabilities expresses the assumption that the output process at time t
11 depends only on the underlying Markov chain at time t .

12

13 **Markov switching linear mixed models**

14

15 Let $Y_{a,t}$ be the observed variable and let $S_{a,t}$ be the non-observable state for individual a
16 ($a = 1, \dots, N$), at time t ($t = 1, \dots, T_a$). The output process $\{Y_{a,t}\}$ of the Markov switching linear
17 mixed model for individual a is related to the state process $\{S_{a,t}\}$, which is a finite-state Markov
18 chain, by a linear mixed model (a linear mixed model can be viewed as an extension of a
19 classical linear model where random effects are added to fixed effects); see Verbeke and
20 Molenberghs (2000). It assumes that the vector of repeated measurements for each individual
21 follows, in each state, a linear regression model where some of the regression parameters are
22 population-specific (i.e. the same for all individuals), whereas other parameters are individual-
23 specific. In our case, the individual status (compared with the average individual) is assumed to
24 be different in each state:

$$\begin{aligned} \text{Given state } S_{a,t} = s_{a,t}, \quad Y_{a,t} &= X_{a,t} \beta_{s_{a,t}} + \tau_{s_{a,t}} \xi_{a,s_{a,t}} + \varepsilon_{a,t}, \\ \xi_{a,s_{a,t}} &\sim N(0,1), \quad \varepsilon_{a,t} | S_{a,t} = s_{a,t} \sim N(0, \sigma_{s_{a,t}}^2). \end{aligned}$$

The Gaussian observation probabilities are defined as follows:

$$b_j(y_{a,t}) = P(Y_{a,t} = y_{a,t} | S_{a,t} = j, \xi_{a,j}) = \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left\{-\frac{(y_{a,t} - X_{a,t} \beta_j - \tau_j \xi_{a,j})^2}{2\sigma_j^2}\right\}.$$

In this definition, $X_{a,t}$ is the Q -dimensional row vector of explanatory variables 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, $\xi_{a,s_{a,t}}$ is the individual a random effect, $\tau_{s_{a,t}}$ is the standard deviation for the random effect and $\sigma_{s_{a,t}}^2$ is the residual variance. The individuals are assumed to be independent. For convenience, random effects are assumed to follow the standard Gaussian distribution. The random effects for an individual a are assumed to be mutually independent ($\text{cov}(\xi_{a,i}, \xi_{a,j}) = 0; i \neq j$). Observations in different states for an individual a are assumed to be conditionally independent given states (for $t \neq t'$, $\text{cov}(Y_{a,t}, Y_{a,t'} | S_{a,1}^{T_a} = s_{a,1}^{T_a}) = 0$ if $s_{a,t} \neq s_{a,t'}$ and $\text{cov}(Y_{a,t}, Y_{a,t'} | S_{a,1}^{T_a} = s_{a,1}^{T_a}) = \tau_{s_{a,t}}^2$ if $s_{a,t} = s_{a,t'}$ where $S_{a,1}^{T_a} = (S_{a,1}, S_{a,2}, \dots, S_{a,T_a})$ denotes the T_a -dimensional vector of non-observable states for individual a).

Semi-Markov switching linear mixed models

In the same manner as for a Markov switching linear mixed model, a semi-Markov switching linear mixed model can be defined, the only difference being the replacement of the underlying Markov chain by an underlying semi-Markov chain.

		state 1	state 2	log-likelihood
No. leaves	SMS-LM	NB(1, 1.56, 0.16)	NB(1, 1.17, 0.11)	-2241.86
		9.2, 7.15	10.27, 9.09	
	MS-LM	NB(1, 1, 0.1)	NB(1, 1, 0.09),	-2242.7
		9.62, 9.1	11, 10.47	
Annual shoot length	SMS-LM	NB(1, 1.91, 0.13)	NB(1, 1.3, 0.16)	-3913.06
		13.63, 9.81	7.83, 6.54	
	MS-LM	NB(1, 1, 0.07)	NB(1, 1, 0.12)	-3915.95
		15.19, 14.67	8.19, 7.67	

Table 1. “Number of leaves” and “annual shoot length” models: Estimated occupancy distributions (parameters, mean and standard deviation in years) for states 1 and 2 for the three-state semi-Markov switching linear model (SMS-LM) and Markov switching linear model (MS-LM).

	p_{11}, λ_1	p_{22}, λ_2	η_3
No. leaves	0.81, 5.34	0.94, 15.59	19.66
Annual shoot length	0.91, 10.64	0.9, 9.72	16.9

Table 2. Markov switching linear mixed models: self-transition probabilities p_{ii} and corresponding mean time λ_i spent in the transient states 1 and 2 (in years), and mean time up to the first occurrence of state 3 η_3 (in years) for the “number of leaves” and the “annual shoot length” models. It should be noted that $\eta_3 < \lambda_1 + \lambda_2$ since a certain proportion of the trees started directly in state 2 (with $\pi_2 = 0.21$ for the “number of leaves” model and $\pi_2 = 0.32$ for the “annual shoot length” model).

		state		
		1	2	3
regression parameters	intercept β_{j1} (s.e.)	3.36 (0.03)	4.47 (0.04)	6.21 (0.12)
	temperature parameter β_{j2} ($^{\circ}\text{C}^{-1}$) (s.e.)	0.08 (0.02)	0.17 (0.03)	0.48 (0.09)
	average temperature effect $\beta_{j2} \times \text{mad}_j(X_2)$	0.09	0.15	0.38
variability decomposition	random variance τ_j^2 (s.e.)	0.41 (0.45)	0.52 (0.39)	0.29 (0.61)
	residual variance σ_j^2 (s.e.)	0.23 (0.01)	0.57 (0.02)	0.7 (0.05)
	total variance Γ_j^2	0.64	1.09	0.99
	proportion of inter-individual heterogeneity	65.08%	47.71%	29.29%
marginal observation distribution (μ_j, Γ_j)		3.36, 0.8	4.56, 1.04	6.55, 0.99

Table 3. Number of leaves per annual shoot: Estimated Markov switching linear mixed model parameters and marginal observation distributions for each state. For each observation linear mixed model the intercept, the regression parameter for the average daily maximum temperature, the average daily maximum temperature effect and the variability decomposition are given. Standard errors are given in parentheses.

		state		
		1	2	3
regression parameters	intercept β_{j1} (cm) (s.e.)	4.01 (0.06)	5.86 (0.14)	12.11 (0.74)
	temperature parameter β_{j2} (cm °C ⁻¹) (s.e.)	0.13 (0.05)	1.53 (0.11)	4.25 (0.6)
	average temperature effect $\beta_{j2} \times \text{mad}_j(X_2)$ (cm)	0.15	1.68	2.61
	cumulative rainfall parameter β_{j3} (cm mm ⁻¹) (s.e.)	$0.5 \cdot 10^{-3}$ ($0.2 \cdot 10^{-3}$)	$4 \cdot 10^{-3}$ ($0.4 \cdot 10^{-3}$)	$5 \cdot 10^{-3}$ ($1.4 \cdot 10^{-3}$)
	average cumulative rainfall effect $\beta_{j3} \times \text{mad}_j(X_3)$ (cm)	0.08	0.54	0.72
variability decomposition	random variance τ_j^2 (s.e.)	0.99 (0.51)	3.82 (0.63)	9.96 (1.71)
	residual variance σ_j^2 (s.e.)	1.3 (0.01)	4.7 (0.21)	29.01 (2.04)
	total variance Γ_j^2	2.29	8.52	38.97
	proportion of inter-individual heterogeneity	43.23%	44.84%	25.56%
	marginal observation distribution (μ_j, Γ_j) (cm)	4.01, 1.51	6.73, 2.92	15.28, 6.24

Table 4. Annual shoot length: Estimated Markov switching linear mixed model parameters and marginal observation distributions for each state. For each observation linear mixed model the intercept, the regression parameter for the average daily maximum temperature, the average daily maximum temperature effect, the regression parameter for cumulative rainfall, the cumulative rainfall effect and the variability decomposition are given. Standard errors are given in parentheses.

profile of state succession	number of leaves per annual shoot model			annual shoot length model		
	unbranched tree	branched tree	total	unbranched tree	branched tree	total
1	25	0	25	41	0	41
1 → 2	54	1	55	27	0	27
1 → 2 → 3	18	10	28	14	13	27
2	12	0	12	18	0	18
2 → 3	7	11	18	16	9	25

Table 5. Number of trees with the same profile of state succession extracted from the optimal segmentations computed using the estimated Markov switching linear mixed models. A profile of state succession is defined as the series of distinct states visited along the sequence regardless the time spent in each state.

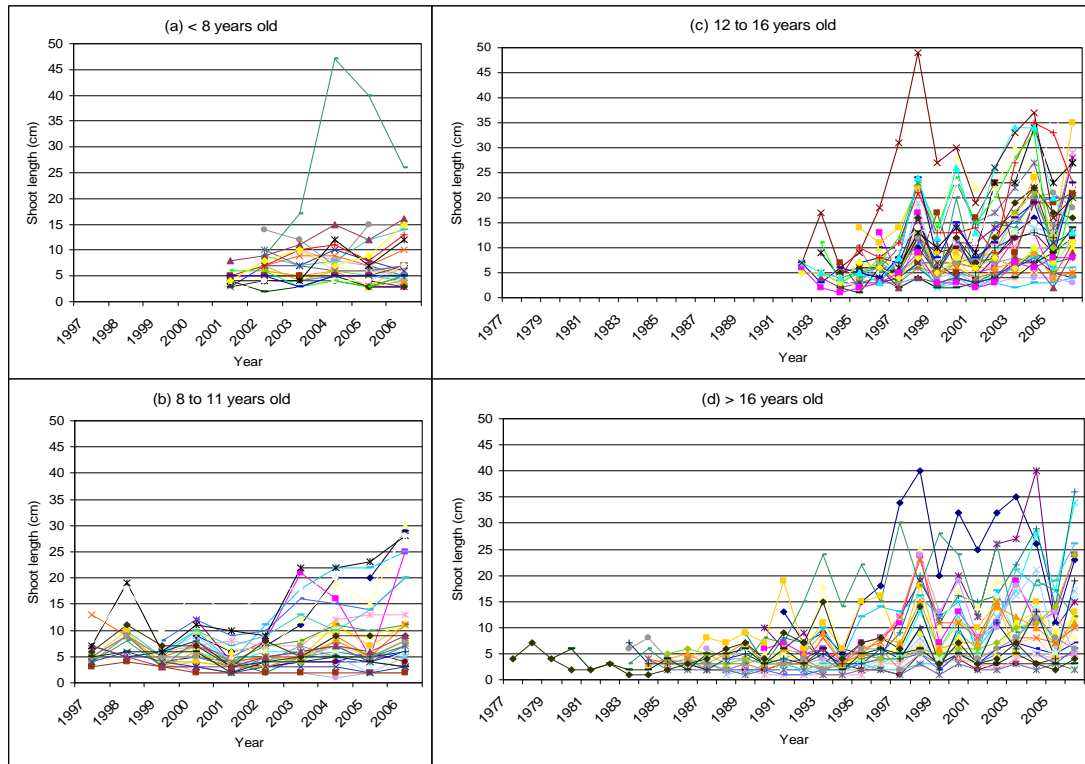


Figure 1. Length of successive annual shoots along walnut main stems (a) less than 8 years old, (b) from 8 to 11 years old, (c) from 12 to 16 years old, (d) more than 16 years old. The very first annual shoot corresponding to the germination year is not shown.

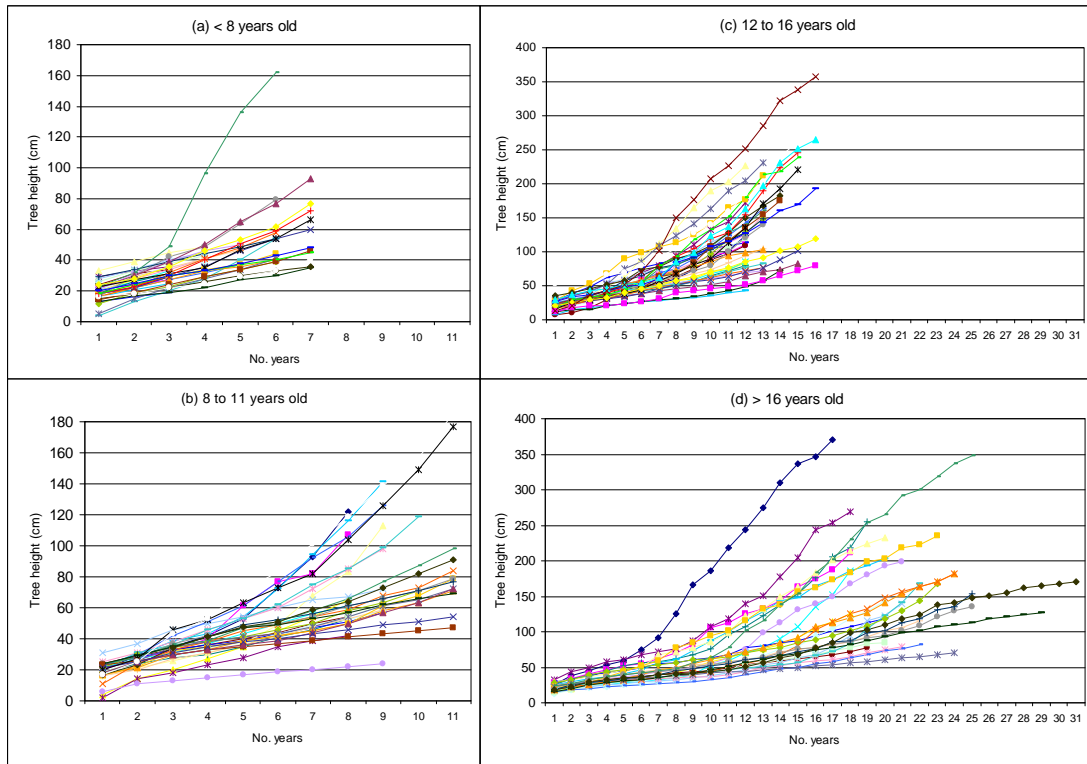


Figure 2. Walnut height (a) less than 8 years old, (b) from 8 to 11 years old, (c) form 12 to 16 years old, (d) more than 16 years old.

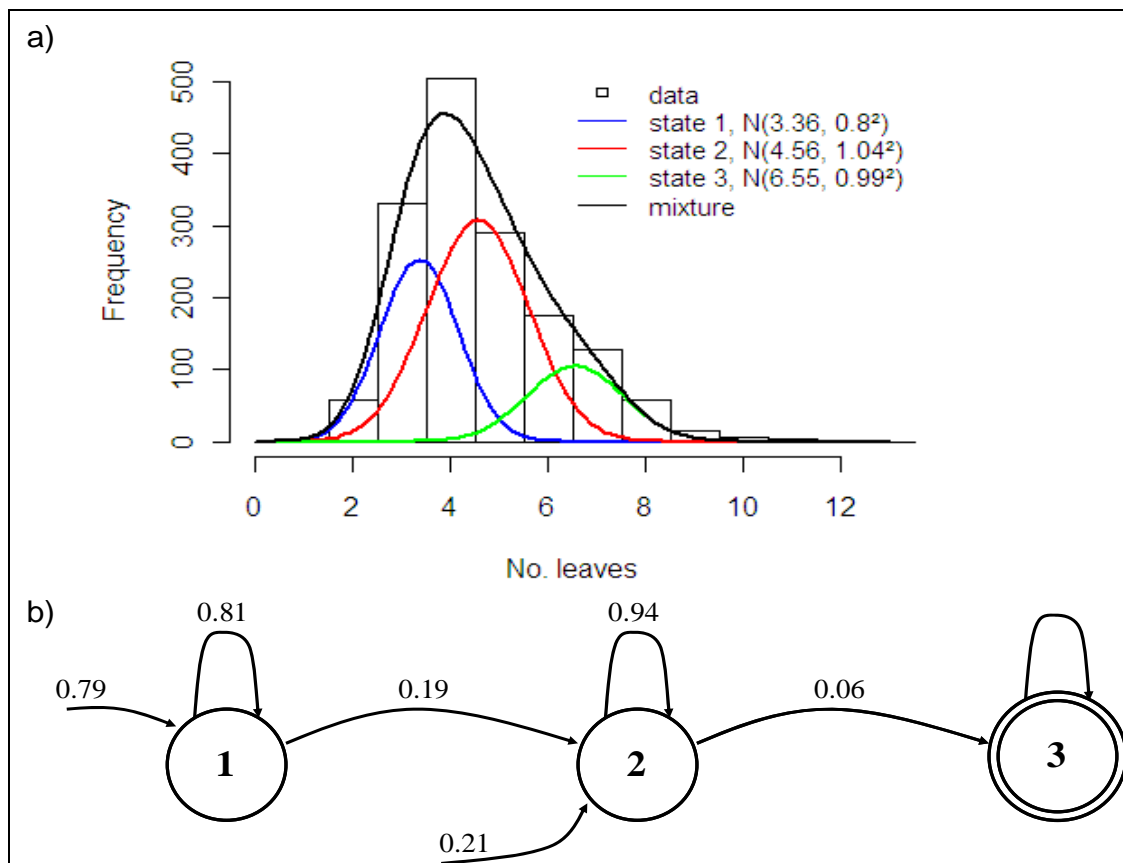


Figure 3. Number of leaves per annual shoot: (a) Mixture of marginal observation distributions. (b) Estimated underlying Markov chain of the Markov switching linear mixed model. Each state is represented by a vertex which is numbered. Vertices representing transient states are edged by a single line while the vertex representing the final absorbing state is edged by a double line. The possible transitions between states are represented by arcs and the attached probabilities are noted nearby. Arcs entering in states indicate initial states. The attached initial probabilities are noted nearby.

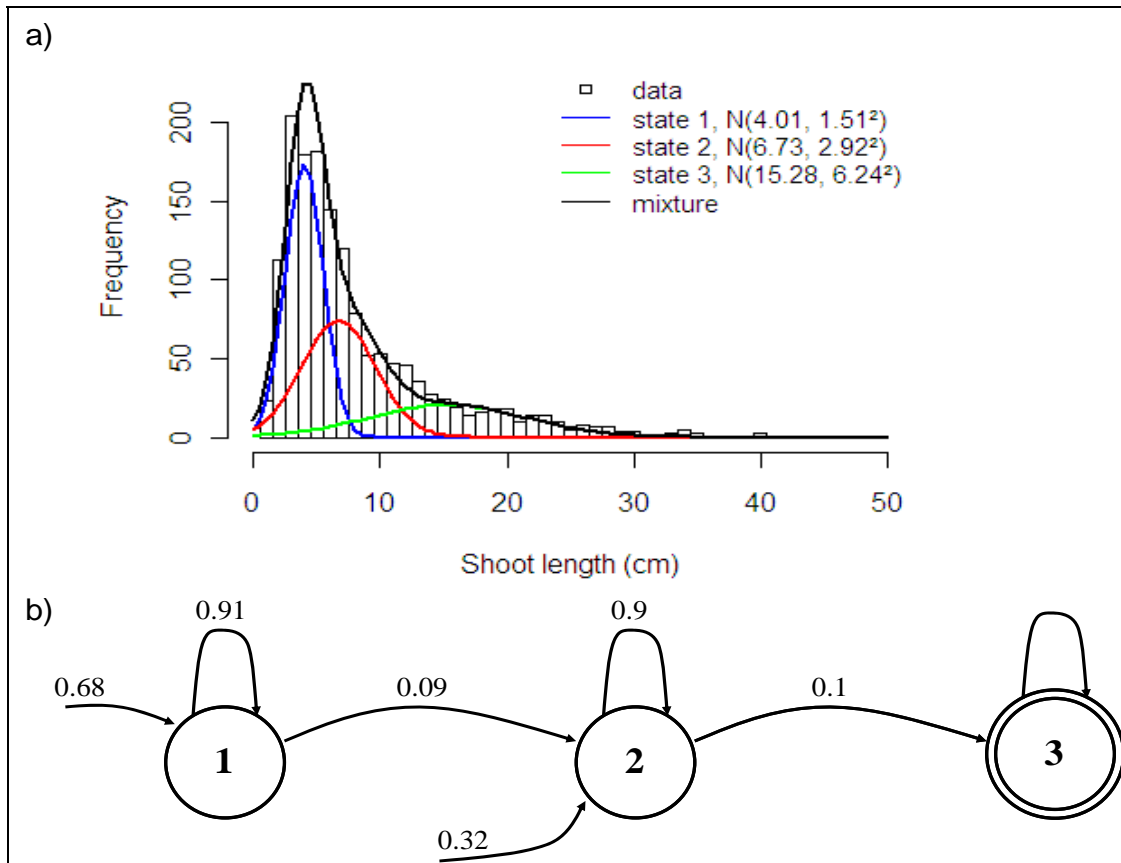


Figure 4. Annual shoot length: (a) Mixture of marginal observation distributions. (b) Estimated underlying Markov chain of the Markov switching linear mixed model. Each state is represented by a vertex which is numbered. Vertices representing transient states are edged by a single line while the vertex representing the final absorbing state is edged by a double line. The possible transitions between states are represented by arcs and the attached probabilities are noted nearby. Arcs entering in states indicate initial states. The attached initial probabilities are noted nearby.

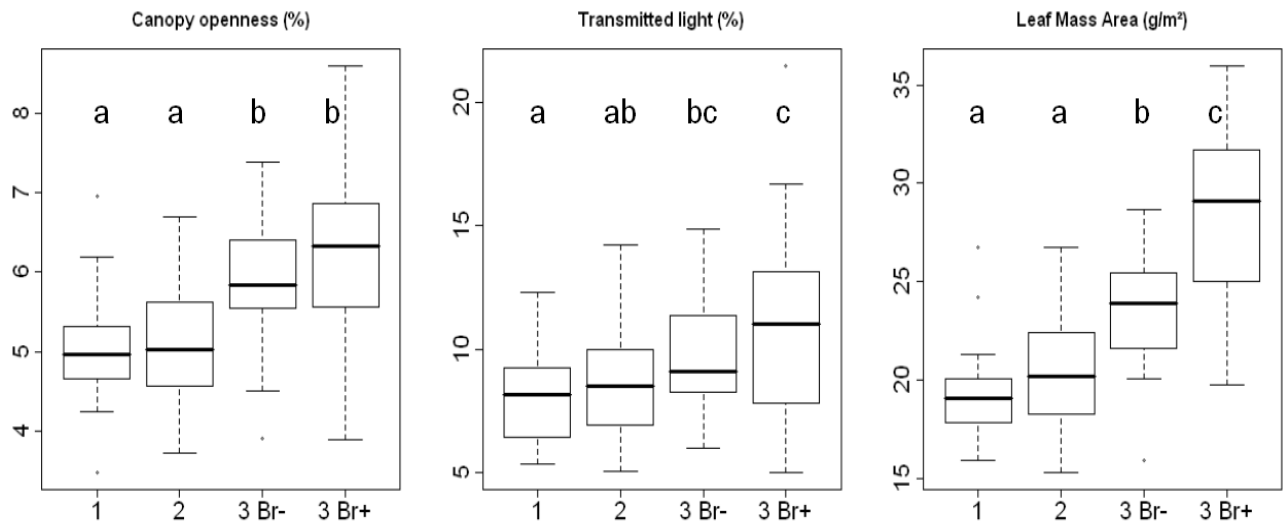


Figure 5. Relationship between light environment and, growth phase and branching status. Box plots for the percentage of canopy openness, the percentage of transmitted light and the LMA values according to growth phase and branching status of trees in 2006 (the last year considered): growth phase 1, growth phase 2, growth phase 3 for unbranched trees (Br-) and branched trees (Br+). Letters indicate homogeneous groups (Tukey test at the 5% level).

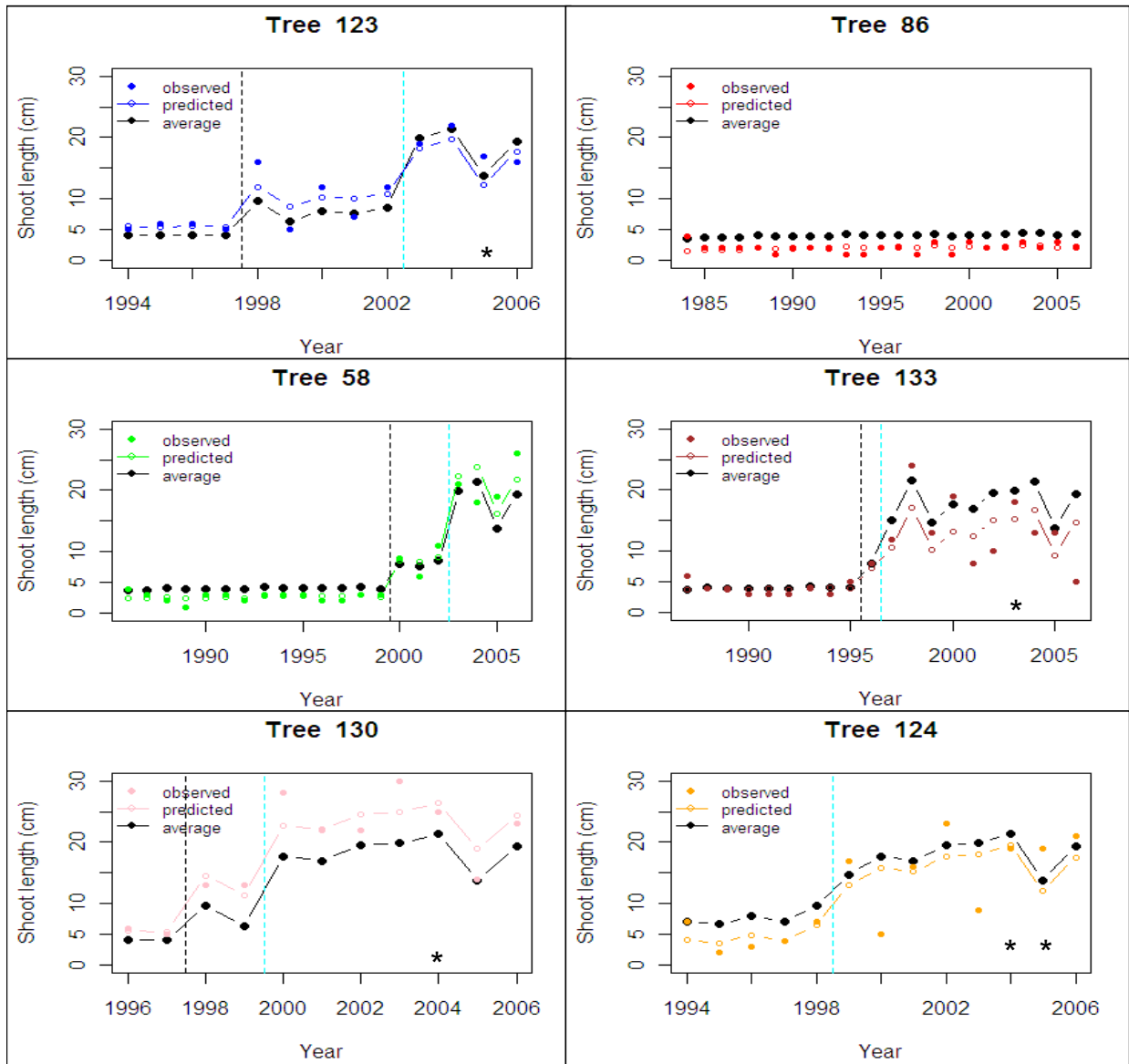
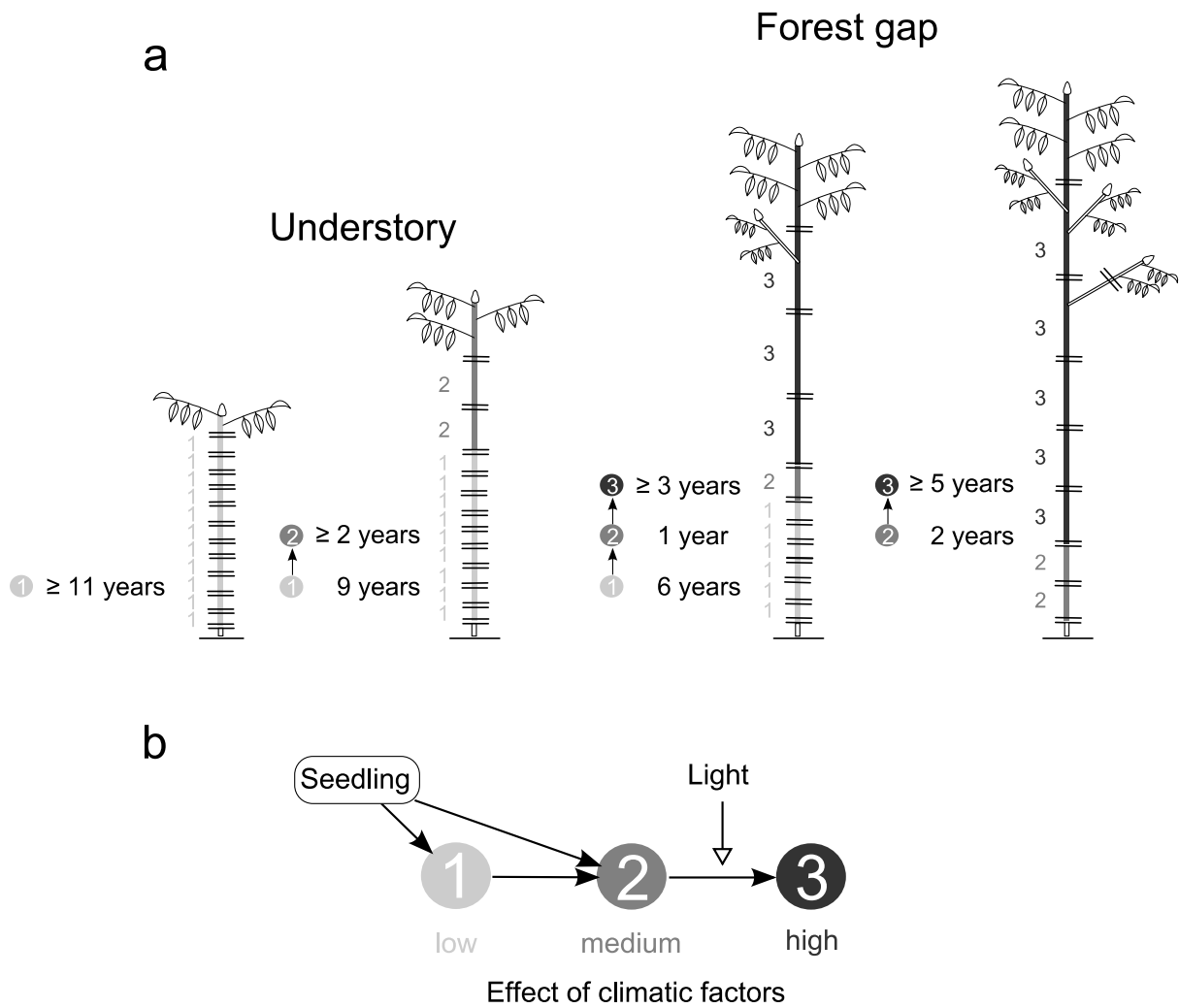


Figure 6. Annual shoot lengths of 6 selected trees: The average tree profiles are represented by black point lines, the predicted tree profiles are represented by color point lines and observed annual shoot lengths are represented by filled circles. The average and predicted tree profiles were computed for each tree on the basis of the optimal segmentation of the observed sequence computed using the estimated Markov switching linear mixed model. The asterisks correspond to years of branching. The black vertical dashed line corresponds to the transition from state 1 to state 2. The cyan vertical dashed line corresponds to the transition from state 2 to state 3.



1
 2 Figure 7. Development of walnut saplings: (a) Four representative individuals: the successive
 3 annual shoots along the main stem are labeled with the growth phases deduced from the
 4 optimal segmentation of the observed sequence computed using the “annual shoot length”
 5 model. The tree development trajectories are summarized nearby. (b) Effects of climatic factors
 6 and individual light environment on successive growth phases.