Math. Model. Nat. Phenom.
Vol. 6, No. 2, 2011, pp. 133-159
DOI: 10.1051/mmnp/20116205

Some Parameter Estimation Issues in Functional-Structural Plant Modelling

P.-H. Cournède[1,2*], V. Letort[1,2], A. Mathieu[3], M.Z. Kang[4], S. Lemaire[5]
S. Trevezas[2], F. Houllier[6], P. de Reffye[7,2]

[1] Ecole Centrale Paris, MAS, Châtenay-Malabry, France
[2] INRIA Saclay - Île-de-France, EPI Digiplante, Orsay, France
[3] AgroParisTech, UMR EGC, Grignon, France
[4] CASIA, LIAMA, Beijing, China
[5] ITB, Paris, France
[6] INRA, UMR AMAP, Montpellier, France
[7] CIRAD, UMR AMAP, Montpellier, France

Abstract. The development of functional-structural plant models has opened interesting perspectives for a better understanding of plant growth as well as for potential applications in breeding or decision aid in farm management. Parameterization of such models is however a difficult issue due to the complexity of the involved biological processes and the interactions between these processes. The estimation of parameters from experimental data by inverse methods is thus a crucial step. This paper presents some results and discussions as first steps towards the construction of a general framework for the parametric estimation of functional-structural plant models. A general family of models of Carbon allocation formalized as dynamic systems serves as the basis for our study. An adaptation of the 2-stage Aitken estimator to this family of model is introduced as well as its numerical implementation, and applied in two different situations: first a morphogenetic model of sugar beet growth with simple plant structure, multi-stage and detailed observations, and second a tree growth model characterized by sparse observations and strong interactions between functioning and organogenesis. The proposed estimation method appears robust, easy to adapt to a wide variety of models, and generally provides a satisfactory goodness-of-fit. However, it does

*Corresponding author. E-mail: paul-henry.cournede@ecp.fr

Article published by EDP Sciences and available at http://www.mmnp-journal.org or http://dx.doi.org/10.1051/mmnp/20116205

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

not allow a proper evaluation of estimation uncertainty. Finally some perspectives opened by the theory of hidden models are discussed.

Key words: functional-structural plant models, carbon allocation, GreenLab, maximum likelihood estimator, Aitken Estimator, hidden models
AMS subject classification: 92B05, 92B15

1. Introduction

To describe plant growth, crop models or process-based models were first developed [31], [5]. They generally aim at predicting yields in fields or greenhouses at a global scale (for example grain yield per square meter). System observation is generally easy and the main issue in parameter estimation concerns identifiability: the number of parameters can be large compared to the dimension of the state / observation vector. [47] recalls the classical estimation methods for crop models. Recent advances in individual plant growth modelling have opened the way to improvements of crop models by describing the functional growth dynamics at organ level, taking into account plant structure and organogenesis. Such models are generally referred to as functional-structural models [57], [23]. However, the complexity of the phenomena involved when modelling the growth of individual plants has sometimes led to the development of very heavy simulation models, capitalizing an important amount of biological knowledge but very difficult to calibrate from experimental data and thus unproper for real applications. Simulated annealing was however used for calibration in [12] resorting to very heavy computations. In this context however, recent efforts have been made to develop the mathematical formulation of functional-structural models as dynamic systems, see [18], [63], [8]. The resulting dynamic systems are non-linear and multi-variable (with very high dimensions of the state-space). Based on this formalism, the objective of this paper is to propose a general framework for the parametric estimation of dynamic models of individual plant growth. Specifically, we will consider morphogenesis models whose state variables are given by the masses or dimensions of plant organs.

Estimation of parameters in dynamic systems has been widely studied in the past decades, we refer for example to [24] or [41]. Most estimation procedures involve adopting a probabilistic framework for model equations by taking into account process and measurement errors. When system observations are regular, very efficient methods based on filtering techniques have been devised. The principle is to treat observed data successively and to improve estimation at each step according to error prediction, see for example extended Kalman filters in [40], recursive least squares in [42] and more recently unscented Kalman filter [32] or particle filtering [17] specifically designed for non-linear systems. These methods have been broadly used and adapted in various domains, including life sciences [54] and modelling for forestry [22].

However, some systems do not allow regular data acquisition, particularly living systems, as measurements are too complex. For the type of plant growth models considered in this study,

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

the variables of interest are generally biomasses (dried or fresh matter). The system output is multivariate and generally corresponds to the masses of some elementary units of the plant (for example all organs, see [27]).

This paper presents some results and discussions as first steps towards the construction of a general framework for the parametric estimation of functional-structural plant models. A general family of models of Carbon allocation formalized as dynamic systems serves as the basis for our study (Section 2.). An adaptation of the 2-stage Aitken estimator (see [58]) to this family of model is introduced (Section 3.) and applied in two different situations: first a morphogenetic model of sugar beet growth with simple plant structure, multi-stage and detailed observations (Section 4.), and second a tree growth model characterized by sparse observations and strong interactions between functioning and organogenesis 5.). Finally, after discussing the properties of the estimation methodology (Section 6.), some perspectives opened by the theory of hidden models are discussed (Section 7.).

2. Dynamic System of Carbon Allocation during Plant Growth

A large number of biophysical processes are studied in the frame of functional-structural plant modelling (*cf.* [23]). In this paper, we concentrate on the carbon budget which can be considered as central regarding morphogenesis. It involves biomass production and allocation [35]. We consider models describing the growth at the organ scale of an individual plant, potentially in a population (field crops, forest stands...). Competition with other individuals for light, water, nutrients can thus be taken into account and affects the computation of biomass production or allocation. Literature is very rich with models of this type, and we refer to [35], [53], [23], [21] for reviews.

2.1. State Variables and State Equations

Organs in plant architecture are characterized by:

- their botanical types: blades, petioles, internodes, flowers, fruits, roots but also potentially sub-categories, for example distinction can be made according to organ physiological age to differentiate organs on the main trunk and secondary branches [4]. Even for complex tree architectures, it is possible to factorize the population of organs to finally get a reasonable total number of categories (as shown by [9]).

- their times of appearance, since the structure is developing.

We will denote by $\mathcal{O}$ the set of botanical types and we let $\{\tau_k\}_{k\in\mathcal{K}}$ denote all the (discrete) times of organ initiations, where K represents an index set. Let $\mathcal{K}_o \subset \mathcal{K}$ denote an index subset corresponding to the times of initiations specifically for organs of type o. For example in maize, there are several times of initiations for blades, while there is only one time of initiation for the cob. If we consider a finite time of growth T, $\forall k \in \mathcal{K}$, $\tau_k \leq T$ and $\mathcal{K}$ is finite. We suppose that the botanical

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

categories are defined in such a way that all organs with the same type and same time of initiation are identical.

For models of carbon allocation at the organ scale, the state variables of interest are thus the biomasses of all types of organs. Let $(o,k) \in \mathcal{O} \times \mathcal{K}_o$ and let $q_{o,k}(t)$ denote the biomass at time t of an organ of type o and time of initiation τ_k. Then the vector of state variables $q(t)$ is given by

$$q(t) = \left(q_{o,k}(t)\right)_{o\in\mathcal{O},k\in\mathcal{K}_o}. \tag{2.1}$$

The state equation is thus of the general form:

$$\frac{dq(t)}{dt} = f\left(q(t), E(t), P, t\right) \tag{2.2}$$

with $E(t)$ a set of environmental variables at time t and P a vector of biophysical or empirical parameters. If the state equation is valid under a wide range of environmental conditions for the same set of parameters P, then we can deduce a strong genetic determinism of these parameters. It was underlined as a great potential interest of plant ecophysiological models in order to develop a predictive capacity that scales from genotype to phenotype [29], [39].

Generally, models of biomass allocation are based on energetic budgets that are not true on a continuous basis, and discrete versions of the state equation (2.2) are derived by considering $(t_n)_{n\in[0;N]}$ the finite sequence of succesive times corresponding to the evolution steps (for example days for crop models or years for forestry models), and we write:

$$q\left(t_{n+1}\right) = q\left(t_n\right) + \int_{t_n}^{t_{n+1}} f\left(q(t), E(t), P, t\right) dt \tag{2.3}$$

2.2. The Example of Source-Sink Models

The state function f may take different forms according to the type of model used (empirical or allometry models, transport resistance ...). We illustrate the formulation of the state equation for the source-sink approach first introduced in [62] and on which are based a large variety of models of carbon allocation (TOMSIM [30], GRAAL [18], GreenLab [66], ECO-Meristem [45] ...). Sources correspond initially to the seed and then to biomass production and reserve remobilization. Sinks are demands for biomass of all living organs [20]. For the sake of simplicity, we consider that all sinks have access to all sources, this hypothesis is that of the common pool of biomass and is shown to be relevant for small plants [14]. However, this hypothesis is not fundamental and could be relaxed.

Let $s_{o,k}(t)$ denote the sink value at time t of an organ of botanical type o and time of initiation τ_k and $N_{o,k}(t)$ the total number of such organs at time t. Of course, if $t < \tau_k$, $s_{o,k}(t) = N_{o,k}(t) = 0$. Under these assumptions (source-sink model with a common pool of reserves for biomass allocation to all living organs) the state equation of carbon allocation is given $\forall (o,k) \in \mathcal{O} \times \mathcal{K}_o$:

$$\begin{cases} \dfrac{dq_{o,k}(t)}{dt} = \dfrac{s_{o,k}(t)}{D(t)} Q(t) \text{ if } t \geq \tau_k \\ \\ q_{o,k}(t) = 0 \text{ if } t \leq \tau_k. \end{cases} \tag{2.4}$$

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

where:

- $D(t)$ is the total demand of the plant at time t (that is to say the sum of all sinks)

$$D(t) = \sum_{o,k} N_{o,k}(t) s_{o,k}(t) \tag{2.5}$$

- $Q(t)$ is the biomass available for allocation (and can be considered as the system input).

$Q(t)$ generally corresponds to biomass accumulation by photosynthesis, but can also result from biomass remobilization. If this phenomenon is taken into account, a negative term will be added in the right hand side of Equation (2.2) to describe the consecutive loss of organ mass.

We denote

$$s(t) = (s_{o,k}(t))_{(o,k)\in\mathcal{O}\times\mathcal{K}_o} \text{ and } N(t) = (N_{o,k}(t))_{(o,k)\in\mathcal{O}\times\mathcal{K}_o}$$

and if v_1 and v_2 are two vectors of the same dimension, $v_1 \cdot v_2$ is their scalar product. Then the state equation 2.2 is given by:

$$\frac{dq(t)}{dt} = \frac{s(t)}{N(t).s(t)} Q(t) \, . \tag{2.6}$$

This equation can serve as the frame for a lot of models, depending on how the variables $N(t)$, $s(t)$ and $Q(t)$ are determined. We refer to [8] who describes the different versions of the GreenLab model in this frame, according to the level of system interactions. The basic version considers:

- $N(t)$ as given (for fixed organogenesis)
- $s(t)$ as empirical functions only depending on some parameters: $s(t, P)$
- $Q(t)$ simply resulting from biomass accumulation and given by ecophysiological laws, for example by Beer's law, as a function of the environmental conditions, green leaf surface area (deduced from leaf biomass) and biophysical parameters (energetic efficiency, Beer's extinction coefficient ...): $Q(q(t), E(t), P)$.

3. Parameter Estimation of Plant Growth Models

3.1. Basic Concepts on the Maximum Likelihood Estimation of Discrete Dynamic Systems

In this section, we recall some useful results on the estimation of discrete dynamic systems. We refer for example to [24] for a comprehensive presentation of the subject.

We consider a dynamic system S described by a discrete time model. Let $(t_n)_{n\in[0;N]}$ be the finite sequence of succesive times corresponding to the evolution steps. We denote by $X_n \in \mathbb{R}^x$ the set of characteristic system variables at t_n, $X_n = X(t_n)$, and we denote by $U_n \in \mathbb{R}^u$, the set of

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

exogeneous variables (entries, controls...) at t_n, and $P \in \mathbb{R}^p$, the vector of model parameters. The type of models addressed in this paper can be written in the following form:

$$X_{n+1} = F_n(X_n, U_n, P) \tag{3.1}$$

with X_0 given.
Let $Y \in \mathbb{R}^y$ be a vector of experimental observations made on the system at one or several times. These observations correspond to model outputs $\tilde{Y}$ deduced from $(X_n)_n$:

$$\tilde{Y} = G((X_n)_n, P) \tag{3.2}$$

If the initial state X_0 and $(U_n)_n$ are known, $\tilde{Y}$ is (implicitly) a function of P. Model errors are represented by the random vector ϵ :

$$\epsilon = Y - \tilde{Y}(P) \ . \tag{3.3}$$

In this article, we choose to use the maximum likelihood estimator since it often has some optimal properties, like asymptotic efficiency (see for example [11]). The conditional likelihood of Y given P, $\Pi(Y|P)$, is the probability density of Y when P is the vector of model parameters. Identification of the system parameters can be performed by finding P^* that maximizes $\Pi(Y|P)$.
We suppose that ϵ is a Gaussian vector of covariance matrix Σ and zero mean. Since the model is deterministic, $\Pi(Y|P,\Sigma) = \Pi(\epsilon)$ and we have:

$$\Pi(Y|P,\Sigma) = [(2\pi)^y \det(\Sigma)]^{-1/2} \exp\left[-\frac{1}{2}\left(Y - \tilde{Y}(P)\right)^T \Sigma^{-1}\left(Y - \tilde{Y}(P)\right)\right] \tag{3.4}$$

If Σ is known, P^* is given by the Gauss-Markov estimator, that is to say by minimizing the quadratic criteria :

$$J(P) = \left(Y - \tilde{Y}(P)\right)^T \Sigma^{-1}\left(Y - \tilde{Y}(P)\right) \ . \tag{3.5}$$

However, the covariance matrix Σ is generally unknown. In such a case, estimators of Σ and P^* can be derived if we have a large number of repetitions of the experimental observations. It is not the type of problem addressed here, since for plants, repetitions cause other problems due to the strong variability between individuals. This issue will be tackled in further studies.

The classical way to overcome this problem is thus to model the error covariance. The simplest possibility is to suppose that output errors are independent and homoscedastic, that is to say $\Sigma = \sigma^2 I$, where I is the identity matrix. The maximum likelihood estimator is thus equivalent to the ordinary least-square estimator. However, this choice is often too limited and other types of model for the error covariance matrix are preferable.

3.2. A 2-stage Aitken Estimator applied to Plant Growth System

The fundamental equation of Carbon allocation (2.3, 2.4) is in the form of a classical discrete dynamic system (3.1). Depending on the particular model chosen, the set of parameters may vary. Some parameters can be directly measured on the field (organ expansion and life spans, mass per unit area...) and the others are to estimate from experimental data.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

3.2.1. Observations

For the type of models considered here, that is to say models of carbon allocation at the organ scale, the state variables of interest are the biomasses of all types of organs, as detailed in section 2.1.:

$$X_n = (q_{o,k}(t_n))_{(o,k)\in\mathcal{O}\times\mathcal{K}_o}\,.$$

Theoretically, organ masses can be measured and measurements are destructive (we cut the organs to weigh them). Therefore, the basic situation gives the observation vector $\tilde{Y}$ as $\tilde{Y} = X_N$, with t_N the experimental time. It raises however two types of problems.

First, in order to take into account the dynamics of plant growth, it would be interesting to consider an observation vector composed of several intermediate observations at different growth stages. Since measurements are destructive, observations should thus be done on different plants, that are supposed representative of the same dynamic system, with the same parameters. This hypothesis is practically never true due to the high variability in plant populations at field level resulting from local environmental variations, genetic variability within a population or variations in seed masses [15]. Since taking into account this variability through an appropriate population growth model is only at its early stage, a practical method is first to define an average plant or an average of plants at each observation time (*cf.* [36]) and to use multi-stage observations based on this average plant.

Second, the experimental process of weighing all plant organs is very heavy (moreover generally, for a better accuracy, dry masses are considered, which also involves oven-drying the organs). It is all the more true when we consider plants with complex architectures and a large number of organs, like trees. Therefore, we may consider sparse or aggregated observations. Such strategy has been used for maize for example, with compartment masses at some stages (that is to say that all organs of the same type o are weighed together) or measuring individual masses for only some of the leaves (one out of five for example) [46]. For trees, some more complex methodologies were introduced by [26], [38]: a detailed description of the trunk (masses of all growth units) and aggregated description of the main branches (for each branch, all organs of the same type o are weighed together).

Finally, the experimental vector may be very irregular and composite. A general way to formalize the corresponding system observation $\tilde{Y}$ is by:

- defining a set of available types of observation functions $(G_i)_{i\in I}$, with I a finite index set, defined on the state variables, and for each i, $G_i : \mathbb{R}^x \to \mathbb{R}^{\alpha_i}$,
- choosing a set of observation times $(t_k^{exp})_{1\leq k\leq K}$,
- for each observation time t_k^{exp}, choosing a subset of the observation functions $(G_i)_i$ characterized by an index subset $I_k \subset I$.

Finally, the system observation vector $\tilde{Y}$ corresponding to the experimental vector Y is given by:

$$\tilde{Y} = \left(\tilde{Y}^{(1)}, \tilde{Y}^{(2)}, \ldots, \tilde{Y}^{(K)}\right) \tag{3.6}$$

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

with

$$\tilde{Y}^{(k)} = \left(G_i\left(X\left(t_k^{exp}\right)\right)\right)_{i\in I_k}$$

3.2.2. Heteroscedasticity

The error covariance matrix is not known and the homoscedastic hypothesis is not a very good option for the GreenLab model since the masses in Y corresponds to different types of organs with potentially very different size orders (*e.g.* petioles of a few grammes in sugar beet versus the whole root system of 250 g, *cf.* [36]).

We thus consider a more general heteroscedastic hypothesis, corresponding to an error model studied in details by [58] and introduced for the GreenLab model by [67]: the y observation are classified into q groups, with each of the y_i error terms in group i has common unknown variance θ_i $(1 \leq i \leq q)$ and errors supposed mutually independent. For the observation vector on plants, each group corresponds to a type of organ. We suppose:

$$\epsilon = Y - \tilde{Y}(P) \sim \mathcal{N}(0, \Sigma) \quad .$$

with Σ a diagonal matrix of rank y:

$$\Sigma = \begin{pmatrix} \theta_1 I_{y_1} & 0 & 0 & \cdots & 0 \\ 0 & \theta_2 I_{y_2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \theta_{q-1} I_{y_{q-1}} & 0 \\ 0 & \cdots & 0 & 0 & \theta_q I_{y_q} \end{pmatrix} \tag{3.7}$$

with I_k, the identity matrix of rank k. Note that it may induce rearranging the data in the $\tilde{Y}$ vector by grouping the data of the same type, resulting in a different arrangement from the one given in (3.6) based on the list of heterogeneous observations.

3.2.3. Two stage Aitken estimator

For such type of sytems, a two stage Aitken estimator can be used (*cf.* [58]). It will be denoted $\hat{P}_{2SA}$. We first find estimates $\hat{\theta}_i$ for all θ_i and then use the Gauss-Markov estimator for P (*cf.* (3.5)) with an estimated covariance matrix $\hat{\Sigma}$:

$$\hat{P}_{2SA} = Argmin_P \left(\left(Y - \tilde{Y}(P)\right)^T \hat{\Sigma}^{-1} \left(Y - \tilde{Y}(P)\right) \right) \quad . \tag{3.8}$$

If $y_i - p \geq 2$, a usual choice for the estimation of θ_i is given by:

$$\hat{\theta}_i = \frac{1}{y_i - p} \left(Y_i - \tilde{Y}_i\left(\hat{P}_{LS,i}\right)\right)^T \left(Y_i - \tilde{Y}_i\left(\hat{P}_{LS,i}\right)\right) \tag{3.9}$$

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

with $\hat{P}_{LS,i}$ the least-square estimator of P on the i-th sub-sample (the i-th group), *i.e.* :

$$\hat{P}_{LS,i} = Argmin_P \left(\left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right) \right)$$

In the linear case, the properties of this estimator are studied by [58].

Remark 1. *The negative log-likelihood denoted by $L(P, \theta)$ is:*

$$L(P,\theta) = \frac{y}{2} 2\pi + \sum_{i=1}^{q} \frac{y_i}{2} ln(\theta_i) + \frac{1}{2} \sum_{i=1}^{q} \frac{1}{\theta_i} \left(\left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right) \right)$$

so that the maximum-likelihood estimators for P and θ (denoted by $\hat{P}$ and $\hat{\theta}$) are:

$$\hat{\theta}_i = \frac{1}{y_i} \left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right), \text{ for } 1 \leq i \leq q, \tag{3.10}$$

and

$$\hat{P} = Argmin_P \left(\sum_{i=1}^{q} \frac{1}{\theta_i} \left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right) \right) . \tag{3.11}$$

$\hat{P}$ is thus solution of the implicit p-dimensional system of equations:

$$\sum_{i=1}^{q} \frac{y_i}{\left(Y_i - \tilde{Y}_i(\hat{P}) \right)^T \left(Y_i - \tilde{Y}_i(\hat{P}) \right)} \left(\frac{\partial \tilde{Y}}{\partial P} \left(\hat{P} \right) \right)^T \left(Y_i - \tilde{Y}_i(\hat{P}) \right) = 0 .$$

Such equation is generally intractable, and we'd rather solve iteratively Equations (3.10, 3.11) starting with

$$\hat{\theta}_i = \frac{1}{y_i} \left(Y_i - \tilde{Y}_i \left(\hat{P}_{LS,i} \right) \right)^T \left(Y_i - \tilde{Y}_i \left(\hat{P}_{LS,i} \right) \right)$$

then deducing $\hat{P}$ and again correcting $\hat{\theta}$ with this new version of $\hat{P}$, until this procedure stabilizes. Numerical tests however showed that there is no significant difference compared with the $2SA$ estimator given by (3.8,3.9).

Remark 2. *In the linear case, that is to say if there exists a matrix $A \in \mathcal{M}_{yp}(\mathbb{R})$ such that:*

$$Y = \tilde{Y}_0 + AP + \epsilon , \tag{3.12}$$

Equation (3.8) can be solved explicitly:

$$\hat{P}_{2SA} = \left(A^T \hat{\Sigma}^{-1} A \right)^{-1} A^T \hat{\Sigma}^{-1} \left(Y - \tilde{Y}_0 \right) , \tag{3.13}$$

Such formulation is interesting in order to solve numerically Equation (3.8) in the non-linear case. An iterative procedure can be derived. Given $\hat{P}_n$, we deduce $\hat{P}_{n+1}$ by a linear approximation near $\hat{P}_n$. We have:

$$\tilde{Y} \left(\hat{P}_n + \beta \right) \approx \tilde{Y} \left(\hat{P}_n \right) + \frac{\partial \tilde{Y}}{\partial P} \left(\hat{P}_n \right) \beta$$

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

and we compute $\hat{P}_{n+1} = \hat{P}_n + \beta_n$ *with* β_n *chosen according to Equation (3.13):*

$$\beta_n = \left(\left(\frac{\partial \tilde{Y}}{\partial P}\left(\hat{P}_n\right) \right)^T \hat{\Sigma}^{-1} \left(\frac{\partial \tilde{Y}}{\partial P}\left(\hat{P}_n\right) \right) \right)^{-1} \left(\frac{\partial \tilde{Y}}{\partial P}\left(\hat{P}_n\right) \right)^T \hat{\Sigma}^{-1} \left(Y - \tilde{Y}\left(\hat{P}_n\right) \right) .$$

This iterative method to solve the minimization problem given in Equation (3.8) is called Gauss-Newton method, cf. [60]. Numerical tests for the estimation of the GreenLab parameters seem to show that this method is more efficient than quasi-Newton minimization methods.

3.2.4. Estimator variance

In the linear case (*cf.* Eq. (3.12)), the covariance of the Gauss-Markov estimator for P is:

$$Cov\left(\hat{P}\right) = \left(A^T \Sigma^{-1} A\right)^{-1} , \tag{3.14}$$

it corresponds to the Cramer-Rao bound since the Gauss-Markov estimator provides the best estimate for P. When Σ is diagonal and in the case of the 2-stage Aitken estimator, that is to say when Σ is estimated for using the Gauss-Markov estimator, it is proved by [58] that to order $\left(\frac{1}{y}\right)$, the variance of $\hat{P}_{2SA}$ is that of the Gauss-Markov estimator in Eq. (3.14). A correction of the variance to order $\left(\frac{1}{y^2}\right)$ is given by [58].

For the non-linear GreenLab model applied to simple plants (sugar beet, maize ...), Equation (3.14) was shown to be a very good approximation for $Var\left(\hat{P}_{2SA}\right)$ by Monte-Carlo simulations. Such result can be explained by the weak level of non-linearity of the system that has been revealed by applying global sensitivity techniques (Sobol's method) in [65]. The level of non-linearity of the model increases for more complex plants, particularly when it presents strong levels of interactions between structure and functioning, as for trees [64].

4. A characteristic application on experimental data of sugar-beet

In this section, we illustrate the estimation method by reporting the results of the parametric identification of the GreenLab model applied to sugar beet (*Radar* variety). This study was conducted at the French research institute for sugar beet (ITB) and the results were presented by [36].

4.1. Model specificities for sugar beet

The time step for the discretization of the fundamental discretized allocation equation (2.6) is chosen as the architectural growth cycle, that is to say as the time necessary for a new phytomer to

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

appear (it generally corresponds to a constant accumulation of temperature above a base temperature [36]). It is given by:

$$q_{n+1} = q_n + s_n \frac{u_n}{N_n \cdot s_n} \ . \tag{4.1}$$

(For any variable α, α_n stands for the variable α at the n^{th} architectural growth cycle.)

- We have only one physiological age, and a mono-stem plant. The number of organs N_n is given, according to the observed phyllochron (with a decay in the leaf rhythm of appearance at full-cover)
- The organs concerned are blades, petioles, and the root system considered as an organ. They are respectively indexed by b, p and r. The sink functions s_o are chosen to only depend on organ's type o and organ's age since initiation u. They are given by parametric β-density functions, modulated by a norm:

$$s_o(u) = \frac{P_o}{N_o} \left(\frac{u}{T_o}\right)^{a_o-1} \left(1 - \frac{u}{T_o}\right)^{b_o-1} \ , \text{ with } N_o = \left(\frac{a_o-1}{a_o+b_o-2}\right)^{a_o-1} \left(\frac{b_o-1}{a_o+b_o-2}\right)^{b_o-1}$$

- The input variable u_n is a feedback function of the state variables:

$$u_n = PAR_n S_p \mu \left(1 - \exp\left(-k\mathcal{A}_n / S_p\right)\right) \tag{4.2}$$

  with

  - PAR_n the photosynthetically active radiation received during growth cycle n,
  - S_p an empirical coefficient corresponding to a characteristic surface (related to the two-dimensional projection of space potentially occupied by the plant onto the x-y plane)
  - μ is an empirical coefficient related to the Radiation Use Efficiency
  - k is the Beer-Lambert extinction coefficient
  - $\mathcal{A}_n$ is the total green leaf area, given by $\mathcal{A}_n = \frac{1}{e} \sum_k \left(N_{b,k}\right)_n \left(q_{b,k}\right)_n$ where a constant mass per unit area e is assumed. Note that here the variables $\left(N_{b,k}\right)_n$ represent the numbers of active blades (non senescent). After a thermal time T_b^f since their initiations, time that may vary with the leaf, leaf blades become senescent and no longer participate to photosynthesis.

4.2. Parametric identification

Some parameters can be directly observed experimentally: $e, T_b, T_p, T_b^f, T_r, k$ and some parameters are fixed $P_b = 1$ (since sinks are relative to each other, one of them has to be fixed for uniqueness) and $P_r = 400$ (since it appears that the fitting quality is not influenced by the root sink provided this one is big enough). The vector of hidden parameters is thus:

$$P = (\mu, S_p, P_p, \alpha_b, \beta_b, \alpha_p, \beta_p, \alpha_r, \beta_r) \ .$$

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

Concerning the experimental data from which the model parameters are estimated, biomass measurements were carried out at seven different stages. At each date, seven individuals plants were selected and the dry mass of every individual organ (blades, petioles and root storage) was measured. The average of the seven plants was constructed.

The numerical values of the parameters as well as their variances are given in [36] and are not recalled here. As an illustration, Figure 1 shows the fitting curves.

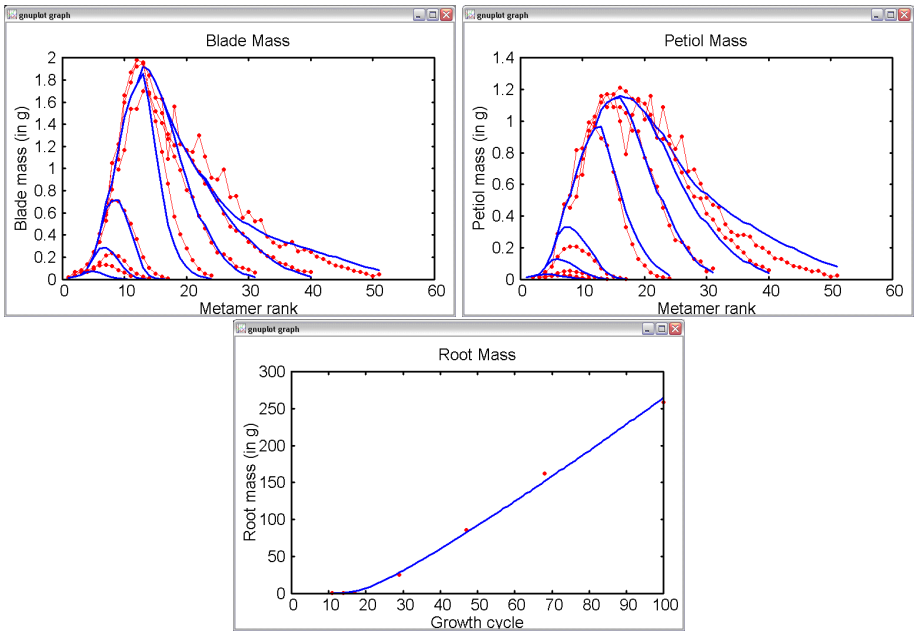


Figure 1: *Fitting results: Average plants at organ level: experimental data (in red) and simulated data (lines) (a) Blade dry masses and (b) petiole dry masses according to phytomer ranks, at 7 different growth stages; (c) root dry mass as a function of the growth cycle (with experimental measures at the same 7 stages)*

5. Application to plants with complex structures: the case of trees

As underlined before, a particularity of mature trees is generally their complex structures. Beside the parameters of the functional model, it is also necessary to estimate the parameters of the organogenesis model. The difficulty arises from the complex interactions between functioning and organogenesis in trees. For some models, the level of interaction is too intricate and the parameter estimation has to be done for the two processes altogether. It is for example the case of the ALMIS [19] or L-Peach [1] models: local conditions like local availability of assimilates and nutrients determine the appearance of new organs. In source-sink models with a common pool of reserves

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

as described in section 2.2., the interaction between functioning and organogenesis is restricted to the number of organs for primary growth, on organ and branch life spans and on secondary growth. Reversely, the functioning has a global retroaction on the appearance of new organs through the ratio of available biomass to demand.

We apply the Aitken estimator to the model of tree growth described by [38], [48] in the GreenLab frame.

5.1. Model specificities for trees

5.1.1. Organogenesis

For most trees grown in temperate zones, growth is rhythmic, *i.e.* there is alternation of periods of activity of the apical meristems and of resting periods. The sequences of times $(t_n)_{n\geq 0}$ will correspond to every active period of the apical meristem. They are characterized by a growth flush and the creation of new *growth units* (Figure 2). A growth unit is composed of a set of phytomers, considered as the elementary botanical units of constant composition (generally an internode, a leaf and potential axillary buds). The concept of physiological age [3] helps differentiate between the different types of axes (and accordingly the organs they are composed of) according to some morphological properties. The main trunk has a physiological age 1 and the oldest physiological age (denoted P) is the ultimate state of differentiation for an axis, it is usually short, without branches. Usually, we need less than 5 physiological ages to describe the axis typology in a given tree.

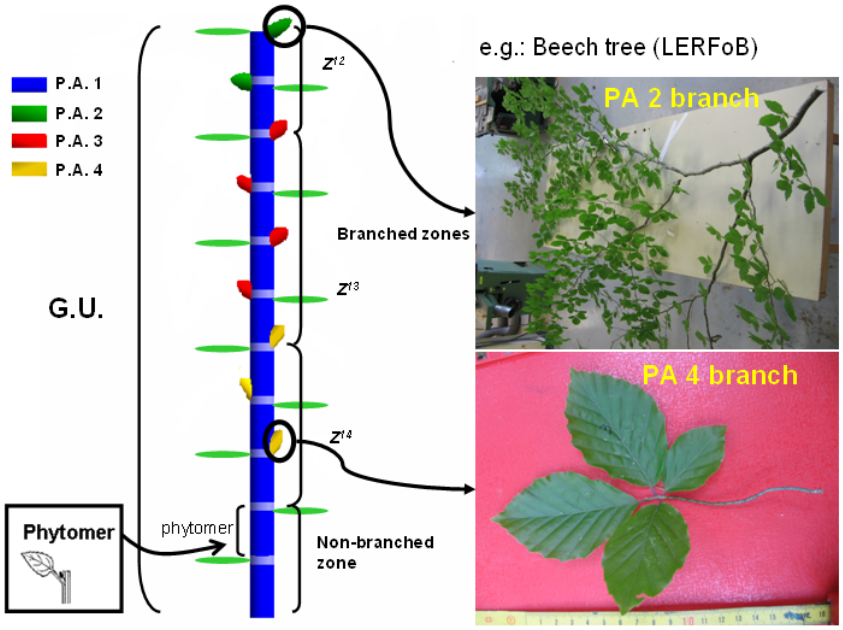


Figure 2: Illustration of a growth unit of beech tree and its constitutive zones Z_{1k}, that are characterized by the physiological ages of the bearing axis (here 1) and that of their lateral buds (k).

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

Growth units are decomposed into zones Z_{ik} that are characterized by the physiological age i of the bearing axis and the physiological age k of the (potential) lateral buds of their phytomers, as introduced by [25]. As described in [9], plant topology can simply be seen as a function of the development sequences $(u_{ik}(t_n))_t$ and $(v_{ik}(t_n))_t$ for all (i,k) such that $1 \leq p \leq q \leq P$, where

- $u_{ik}(t_n)$ gives the number of phytomers of physiological age i with axillary buds of physiological age k in growth units appearing at time t_n (n^{th} growth cycle),
- $v_{ik}(t_n)$ gives the number of growth units of physiological age k borne by axes of physiological i appearing at time t_n (n^{th} growth cycle).

Note that these variables are generally stochastic for trees as detailed in [33]. Different buds of the same type in the tree may have different productions. Usually, characterizing the average organogenesis behavior is sufficient to compute plant functioning. The number of organs of all types $N(t_n) = (N_{o,k}(t_n))_{(o,k)\in\mathcal{O}\times\mathcal{K}_o}$ can be deduced from all the sequences $(u_{ik}(t_n))_n$ and $(v_{ik}(t_n))_n$ as detailed in [9] (or in [33], [43] for the stochastic case).

A fundamental property of trees is their plasticity in response to changes in their growth conditions. Plants are seen as self-regulating systems that can react to exogenous (e.g. environmental stresses that reduce plant production; organ or branch pruning that reduce plant demand) or endogenous (e.g. apparition of new branches or fruits that increase demand) influences. In the GreenLab model, this translates into an index of tree vigor, defined as the ratio of available biomass to total demand (Q/D) [48]. This ratio is representative of the level of trophic competition inside the plant. Thus it allows simulating as well the ontogenetic changes in plant topology throughout its growth phases (progressive set up of architectural units) as architectural plasticity in response to environmental changes. This ratio controls the appearance and longevity of new organs (organogenesis) and branches, thus updating the potential topology at each growth cycle. The development sequences $(u_{ik}(t_n))_n$ and $(v_{ik}(t_n))_n$ can be modeled as functions of this ratio. For instance, the total number of new axes $v_{ik}(t_n)$ actually appearing on zones Z_{ik} in the whole tree depends on the number of existing positions potentially bearing that kind of axes, $N_{ik}(t_{n-1})$, and on the value of Q/D at time t_{n-1}. [38] proposed the following equations to describe the average numbers of phytomers and branches appearing at each growth cycle:

$$\begin{cases} u_{ik}(t_n) = \left\lfloor A^1_{ik} + A^2_{ik} \cdot \frac{Q(t_{n-1})}{D(t_{n-1})} \right\rfloor \\ v_{ik}(t_n) = \left\lfloor N^{ik}(t_{n-1}) \cdot \left(B^1_{ik} + B^2_{ik} \cdot \frac{Q(t_{n-1})}{D(t_{n-1})} \right) \right\rfloor \end{cases} \tag{5.1}$$

where for a real number x, $\lfloor x \rfloor$ represents its rounded value and A^1_{ik}, A^2_{ik}, B^1_{ik}, B^2_{ik} are model parameters.

5.1.2. Ecophysiological model

For trees, the biomass production $Q(t_n)$ can be expressed using the same Beer's law as in Equation (4.2). A difference to take into account the important changes in plant architecture is to consider

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

the parameter S_p (related to the crown projection surface and important to compute self-shading as well as competition with other plants) as an increasing function of tree leaf area ([10]):

$$S_p\left(t_n\right) = S_{p0}\left(\frac{\mathcal{A}\left(t_n\right)}{S_{p0}}\right)^{\alpha}$$

The production function thus involves 3 hidden parameters: the radiation use efficiency μ, S_{p0} and α.

Internode growth can be decomposed into two phases: primary growth which consists in stem elongation and secondary growth, or cambial growth, which consists in diameter increase by successive ring increments (annual increments for trees in temperate zones). Trees are characterized by the predominance of the secondary growth as a factor determining internode mass and size.

In the GreenLab model, consistently with the hypothesis of a common pool of reserves for biomass allocation, there is first a global step of allocation of biomass dedicated to primary and secondary growths respectively at the whole plant level. It corresponds to the hypothesis that plants have a global allocation strategy which generally favors primary growth (new growth units). Secondary growth is mostly used to stock biomass reserves. We denote by $\alpha(t)$, the proportion of biomass allocated to primary growth. After this first step, $\alpha(t)Q(t)$ is distributed among the primary sinks and $\left(1-\alpha(t)\right)Q(t)$ among all the secondary sinks. Equation (2.6) becomes:

$$q_{o,k}(t+1) = q_{o,k}(t_n) + \left(\frac{\alpha(t_n)s_{o,k}(t_n)}{D_1(t_n)} + \frac{\left(1-\alpha(t_n)\right)r_{o,k}(t_n)}{D_2(t_n)}\right)Q(t_n) \tag{5.2}$$

with $s_{o,k}(t_n)$ corresponding to primary sink and $r_{o,k}(t_n)$ to secondary sink (for cambial growth), and

$$D_1(t_n) = \sum_{o,k} N_{o,k}(t_n)s_{o,k}(t_n)\ , \tag{5.3}$$

and

$$D_2(t_n) = \sum_{o,k} N_{o,k}(t_n)r_{o,k}(t_n)\ . \tag{5.4}$$

For most temperate trees, stem elongation and leaf expansion can be considered as immediate. Therefore, primary sinks $s_{o,k}(t_n)$ can be deduced from direct measurements (the ratios of organ weights give directly the allocation coefficients) [26].

There are numerous evidences in the literature that radial growth is more sensitive to changes in growth conditions than primary growth: for instance, trees suffering from high competition can produce long primary shoots but usually narrow rings [50]. Cambial growth can play a buffer role: trees invest more in secondary growth if conditions are favorable. This translates, again, into the index of tree vigor, defined as the ratio of available biomass to total demand (Q/D) [48] and [38] proposed to choose

$$\alpha(t_n) = \frac{1}{1+\frac{1}{P_r\cdot\left(\frac{Q(t_n)}{D(t_n)}\right)^{\gamma}}} \tag{5.5}$$

where P_r and γ are empirical parameters to estimate.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

The formulation of secondary sink has to take into account the physiological age of the internode, as well as its position in plant structure, since its influence on the cambial growth of each phytomer has long been highlighted ([2]). We refer to [38] for a detailed description of the equations giving the secondary sinks. Finally, their expressions involve the parameters $(R_k)_{1\leq k\leq P}$, corresponding to coefficients modulating allocation according to the physiological ages k of branches, and an empirical coefficient λ expressing the balance between uniform allocation (common pool) and topoly-driven allocation (Pressler law, [2]).

5.2. Adaptation of the estimation method

Apart from the parameters that can be measured directly on trees, the basic formulation of the tree ecophysiological model describing source-sink relationships would thus involve the following vector of hidden parameters that need to be estimated by model inversion:

$$P_1 = \left(\mu, S_{p0}, \alpha, P_r, \gamma, (R_k)_{1\leq k\leq P}, \lambda\right).$$

If plant topology is given (that is to say if the development sequences $u_{ik}\,(t_n)_n$, $v_{ik}\,(t_n)_n$ are known for all $1 \leq i \leq k \leq P$), the model outputs are continuous functions of the parameters and the 2-stage Aitken estimator can be applied straightforwardly. Regarding topological parameters, we need to estimate the parameters driving the retroaction of functioning on organogenesis

$$P_2 = \left(A^1_{ik}, A^2_{ik}, B^1_{ik}, B^2_{ik}\right)_{1\leq i\leq k\leq P}.$$

Equations (5.1) include transformations from real values to integers. Hence, they do not generate continuous variations of the model outputs and the Gauss-Newton method described in Section 3.2.3. can no longer be applied. We need to resort to heuristic methods to minimize the cost function (3.8). We chose the simulated annealing algorithm [34].

Owing to the relative simplicity of functional-structural interactions in the GreenLab model, it is possible to perform the estimation of topological and functional parameters in a recursive loop with separated steps. In a first step, the topology is set as observed for each tree. It sets the frame allowing the functioning parameters to be estimated, using a continuous method. Once these functional parameters are estimated, topological data (e.g. number of branches or organs of given physiological age classes) are added into the target data and the topological parameters are set free to vary. Several trials are usually needed to ensure convergence to a reasonable solution with the simulated annealing.

5.3. Application to beech tree

5.3.1. Measurement procedure and data

Two common beech trees (*Fagus sylvatica* L.) grown in understorey were measured in May 2006 from the natural stand of Champenoux, near Nancy (North-Eastern part of France). Tree 1 was 21 year-old and Tree 2 was 46 year-old.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

The topological development of beech tree follows the botanical rules of the Troll architectural model [28]. Following [59], four physiological ages were distinguished: short shoots bearing no branches (physiological age (PA) 4), long shoots bearing only short shoots (PA 3), long ramified shoots (PA 2) and trunk (PA 1). Fig. 2 and 3 illustrate the potential compositions of beech growth units. Samples were taken at random locations in tree crowns in order to estimate the mass per

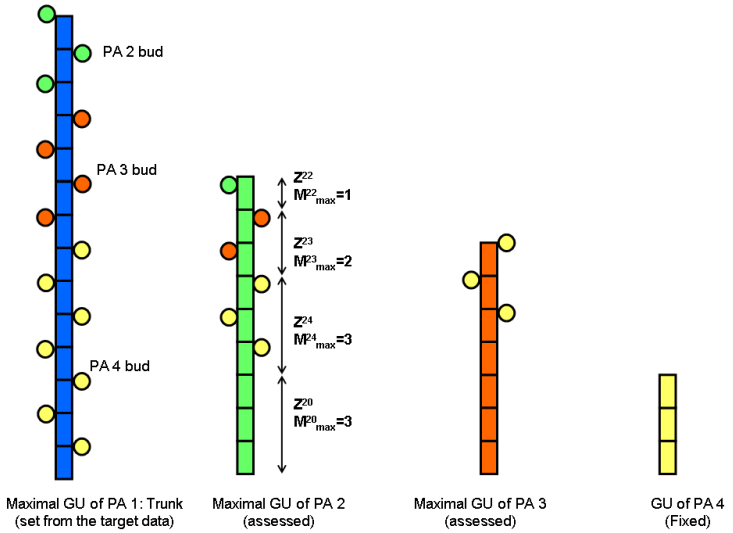


Figure 3: Potential topology for growth units of the four physiological ages. The topology of growth unit on the main stem (PA 1) is set directly from experimental measurements.

unit area, allometric ratios for biomass partitioning (sinks), and allometric relationships linking internode dimensions to their masses.

For mature trees, observation of the state vector $q(t)$ (*cf.* Equation (2.1)) would necessitate prohibitive experimental load. Therefore, some particular choices were made to define the experimental vector $Y(t)$ from which the parameters could be estimated: complete trunk profile (fresh mass, mean diameter and length of each growth unit, plus annual ring width at regular intervals) and compartment biomasses (cumulated biomass of all branch internodes, cumulated biomass of all leaves) for each branch of physiological age 2.

5.3.2. Results of the estimation process

These results are presented in details in [38]. In this section, we underline some specific points characteristic of the estimation process for trees.

As no quantitative topological information was recorded concerning the branching structure of order 2 branches, the topological parameters were fitted on aggregated data of biomass compartments only. The model parameters were estimated to create an average structure with a global demand at each growth cycle similar to the real one. The average simulated weights of blade and internode were compared to the measured data to ensure that the total number of metamers is globally coherent with the experimental observations.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

As no measurement was done concerning the environment factor, it was assessed to be constant during the growth period but specific to the local environment of each tree. It was, therefore, considered as a parameter of the model (instead of a control factor) and its value was estimated from experimental data. It is an indicator of the global level of environmental pressure on tree growth and it is the only parameter that was allowed to vary independently for Tree 1 and Tree 2. For every other parameter, a unique value was estimated from the data of both trees simultaneously. A smaller value of the environment was found for Tree 1 (5.6 vs 10) which was consistent with a higher mass per unit area for Tree 2, usually characteristic of more favorable light conditions.

Fitting results for the trunks are presented in Figure 4. Although data from tree 1 and tree 2 are in very different ranges, they could be fitted in parallel with satisfactory success.

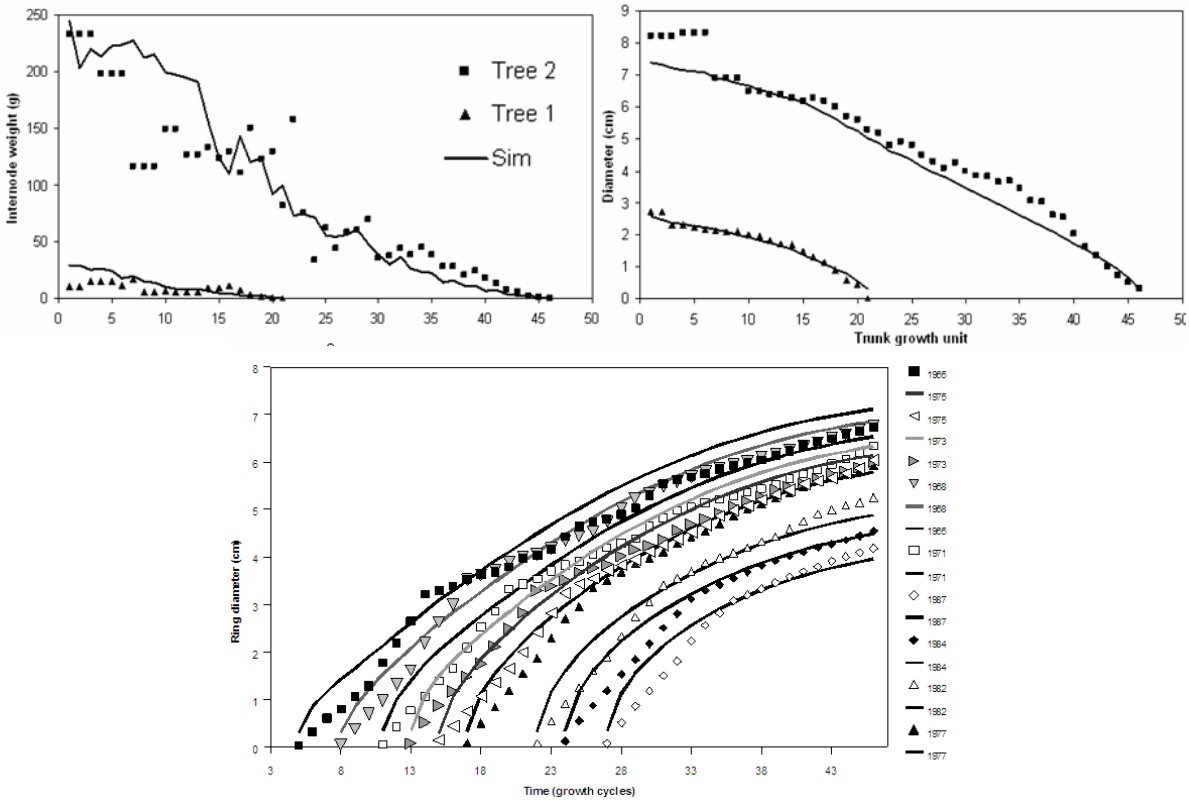


Figure 4: Fitting results: observed and simulated growth units' weights and diameters on the trunk; ring profile for tree 2. Series represent the successive ring diameters of growth units according to their year of appearance.

The experimental vector for tree 2 also included sequences of ring widths for 12 growth units located at different heights near the bottom of the trunk (Figure 4). There is a slight over-estimation of internode pith mass but the global trends are respected. The accuracy of simulation of ring diameters is particularly important as it gives access to information about the past growth of the tree. Thus, not only the last stage but the complete source-sink evolution in time is likely to be correct in the simulated tree.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

Concerning topological parameters, several optima are possible, in particular because the equations like 5.1 include transformations from real values to integers. Therefore, intervals are given for these values [38].

These results illustrate that a simulated tree with a simplified structure can mimic the morphological plasticity of a real tree when the branching structure is controlled by a dynamic feedback of its trophic state on its organogenesis processes. It was also a first proof-of-feasibility of the fitting procedure presented above.

6. Discussion

The aim of this article was to present a framework for the parameter estimation of dynamic systems of carbon allocation in the context of functional-structural plant growth modelling. We believe that the methodology presented is quite generic and could be applied for a large variety of model implementations. It was illustrated on two very different growth models for sugar beet and beech tree involving different levels of model complexity: long organ expansion for sugar beet, strong levels of interactions between organogenesis and functioning, secondary growth for beech tree. Both models however correspond to derivations of the GreenLab generic modelling frame, which has led to extensive variations, for plants as different as maize [27], arabidopsis [7], grapevine [52], pine [26] or cucumber [49], and for which model parameterization was obtained through this methodology, adapted to each particular case. It is proved to be very robust.

If the goodness of fit is in all cases satisfactory, the statistical analysis of the estimation results is not properly done yet, and it makes it difficult to develop a predictive capacity for the model, by controlling the uncertainty of prediction. As a matter of fact, the variances of the estimates are given under the assumption that the error model chosen is the good one and that the model is linear. As shown by [65], the linearity hypothesis might be acceptable in some cases, but the hypothesis of Σ diagonal, with masses of the same organ type having common variance θ_i ($1 \leq i \leq q$) and errors supposed mutually independent, *cf.* (3.7), is quite gross. It corresponds to using only the deterministic part of the dynamical model, which is really restrictive since in dynamic models, errors propagate (see an example on plant growth models in [56]). Current studies aim at improving estimation by modelling the dynamics of the covariance matrix of the model noise.

Moreover, even though the system of study is an individual plant, it belongs to a population of individuals and the results of interest regarding parameter estimation are those that characterize the whole population. Particularly, as underlined by [29], [39], model parameters should be characteristic of a given genotype in order to help bridge the gap between genotype and phenotype for applications of plant growth models in plant breeding. Therefore, an important issue concerns how to combine observations on different individuals that are characterized by strong variability in the estimation process. It may concern local environmental conditions, initial conditions or even genetic variability within the same genotype [15]. So far, the methodology was purely applied on a hypothetic 'average plant', built for example in the sugar beet case as the average of several plants, which was possible owing to the simple structure of sugar beet plants. In the test case on beech trees, we were able to estimate a difference in the environmental effect for the two trees.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

However, as explained above, in terms of retroaction of organogenesis we were only able to assess an average behavior that led to the same global source-sink dynamics at the whole plant level.

In case of complex stochastic organogenesis, the method presented in Section 5.2. should be adapted. The major problem arises when flushes of growth units are not synchronized, especially for tropical trees [13], and it is thus impossible to deduce ages of growth units by direct observations. In the estimation process proposed in Section 5.2., it is thus no longer possible to fix an average topology in order to first estimate the functional parameters, before estimating the parameters of retroaction of functioning on topology. Two different ways are being explored: first by estimating in the first place the parameters of the stochastic organogenesis model from observations on the distributions of the numbers of organs or of specific structural patterns in plants ([44]), or by a combined approach allowing to estimate jointly the functional parameters and organogenesis parameters [61] from architectural data when the stochastic model has few parameters.

Another crucial issue for the application of plant growth models is the simplification of the experimental process. For example, if breeders want to characterize a whole set of genotypes by their specific model parameters, they probably can not afford the heavy experiments that were used to estimate the sugar beet parameters in the example of Section 4.. It would thus be important to optimize the experimental protocol, which implies being able to characterize the uncertainty in parameter estimation according to the level of experimental details. Likewise, as underlined in Section 5.3.1., the same problem occurs for mature trees since the number of organs is prohibitive for measurements. Sampling and aggregation are the keys to experimental simplifications, and several tries were performed, see [37] for trees or [46] for maize, without however proposing a systematic approach of uncertainty evaluation.

For a better control of model uncertainty, it is necessary to improve the statistical framework for modelling. For this reason, we are currently exploring the implementation of functional-structural plant models as hidden models, and the perspectives it opens. The closing section of this paper is thus more like an introduction to current works, and the problems they may resolve.

7. A promising Perspective: the Frame of Hidden Models

Instead of considering the state variables as the vector of organ masses varying with time, the observed organ masses are organized into ordered sequences, in which the ordering is taken with respect to the growth cycle of organ appearance. Such sequences lead to an equivalent description of the dynamic system in the framework of hidden (latent variable) models (*cf.* [6]): model equations can be interpreted as a bidimensional stochastic process, which describes the evolution of system variables.

First, we describe the case where the observation vector consists of measurements on a single plant. More precisely, let $\boldsymbol{Y} := \{Y_n\}_{n\in\mathbb{N}^*}$ be the sequence of observed organ biomasses, ordered as described above with n the growth cycle of appearance. Each Y_n is generally a multidimensional vector, that can be divided into subvectors $Y_{o,n}$, each one corresponding to a different type of organ. Let also $\boldsymbol{Q} := \{Q_n\}_{n\in\mathbb{N}}$ be the sequence of successively produced biomasses Q_n, at each growth cycle n, where Q_0 corresponds to the seed mass. Q_n is considered to be distributed to the different

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

organs at cycle $n+1$. The following assumptions describe generally the representation of the model in this framework, and the dependence relations among system variables:

1. $\boldsymbol{Y}$ is the observed sequence, where $\boldsymbol{Y}$ given $\boldsymbol{Q}$, forms a sequence of conditionally independent random variables, and each subvector $Y_{o,n}$ of Y_n, given $\boldsymbol{Q}$, depends only on $(Q_{n-1}, Q_n, \ldots, Q_{n+T_o-2}) \stackrel{\text{def}}{=} Q_{n-1}^{n+T_o-2}$, where T_o is the characteristic expansion time for each organ of type o. Let $T \stackrel{\text{def}}{=} \max_o T_o$. The vector Q_{n-1}^{n+T-2}, corresponds to the vector of contributing biomasses to the expansion of the organs of Y_n. If we denote by $G_n(Q_{n-1}^{n+T_o-2}, P)$ the empirical function that models its expected value (its exact value, when there is no error), then the vector Y_n given Q_{n-1}^{n+T-2} follows $\mathcal{N}\left(G_n(Q_{n-1}^{n+T-2}, P), \Sigma_Y\right)$, where Σ_Y is an appropriate covariance matrix, not necessarily diagonal, that takes into account the correlation structure of the model.

2. $\boldsymbol{Q}$ is the hidden sequence, and the kind of dependence among the random variables Q_n determines the nature of the hidden model. We present here the case when Q_n given the vector of previous biomasses $(Q_0, Q_1, \ldots, Q_{n-1})$ depends only on Q_{n-1}. In this case, we are in the framework of a Hidden Markov model. If we denote by $R_{n-1}(Q_{n-1}, P)$ the recurrent relation that expresses the corresponding sequence of expected values, then Q_n given Q_{n-1} follows $\mathcal{N}\big(R_{n-1}(Q_{n-1}, P), V_{n-1}(Q_{n-1}, \theta)\big)$, where $V_{n-1}(Q_{n-1}, \theta)$ is an empirical function that models the variance of Q_n around its expected value $R_{n-1}(Q_{n-1}, P)$, and θ an appropriate parameter that can contain or not the parameter P.

Statistical estimation for plant growth models that can be represented in this framework, can be tackled with tools borrowed from the theory of Hidden Markov models (*cf.* [68]). One popular solution is to perform parametric identification via maximum likelihood estimation. The observed vector (after destruction of the plant) consists of organ masses of mature organs (those that have completed their expansion), and of immature organs (those that have not completed their expansion due to the censoring effects). The associated likelihood function cannot be computed in a closed form. In order to approximate the likelihood function, there are at least two different approaches based on iterative optimization techniques. When the state space of the hidden chain is finite, one method is based on the expectation-maximization (EM) algorithm (see [16] and [55]) and the other on gradient type methods (see [68] and the references therein), where direct numerical maximization is implemented. Nevertheless, in our case the state space is continuous and no exact implementation of the EM algorithm or the gradient methods is feasible.

Therefore, simulation based methods are in progress in order to implement proper stochastic versions of the EM algorithm and stochastic gradient methods for state and parameter estimation. In this direction, the class of sequential Monte-Carlo (or particle filter) algorithms, which can be used for maximum likelihood estimation (see [17]), seems particularly adapted to our case. When the complete likelihood function belongs to the exponential family, the EM algorithm is usually well suited and a set of sufficient statistics simplifies the maximization step (see [51]). In our case, the model parameterization makes the maximization step more complicated, and thus stochastic gradient as well as stochastic EM algorithms will be developed and compared.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

When observations of different plants at different times are considered (see Equation 3.6), individual observations can be combined by assuming that they correspond to independent realizations of the same stochastic process. Generally, the censoring time will be different (since plants are not necessarily measured at the same date) and based on the independence assumption, the likelihood function in the new setting model is the sum of the individual likelihood contributions, for each of the observed vectors. A combination of the independence assumption (among observations of different plants) and of the special structure of each individual observation vector allows the extension of the estimation method to this case.

This framework is under implementation (first for unbranched plants) and open promising perspectives in order to control estimation uncertainty.

Acknowledgements

We are grateful to our research colleagues in agronomy and forestry who are crucial partners for this work, in both building models and providing experimental data, specifically ITB (Fabienne Maupas) for sugar beet data and LerFob (Thiéry Constant) for beech tree.

References

[1] M.T. Allen, P. Prusinkiewicz, T.M. Dejong. *Using L-systems for modeling source-sink interactions, architecture and physiology of growing trees, the L-peach model*. New Phytologist, 166 (2005), 869–880.

[2] E. Assmann. The principles of forest yield study. Perfamon Press Ltd., Oxford, 1970.

[3] D. Barthélémy, Y. Caraglio, E. Costes. *Architecture, gradients morphogénétiques et âge physiologique chez les végétaux*. Modélisation et simulation de l'architecture des végétaux, (J. Bouchon, ed.), Sciences Update, INRA, 1997, 89–136.

[4] D. Barthélémy, Y. Caraglio. *Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny*. Annals of Botany, 99 (2007), No. 3, 375–407.

[5] N. Brisson, C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, F. Bussière, Y.M. Cabidoche, P. Cellier, P. Debaeke, J.P. Gaudillère, C. Hénault, F. Maraux, B. Seguin, H. Sinoquet. *An overview of the crop model STICS*. European Journal of Agronomy, 18 (2003), 309–332.

[6] O. Cappé, E. Moulines, T. Rydén. Inference in hidden Markov models, Springer, New York, 2005.

[7] A. Christophe, V. Letort, I. Hummel, P.-H. Cournède, P. de Reffye, J. Lecoeur. *A model based analysis of the dynamics of carbon balance at the whole plant level in Arabidopsis Thaliana*. Functional Plant Biology, 35(11) (2008), 1147–1162.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

[8] P.-H. Cournède. Dynamic system of plant growth. HDR Thesis, University of Montpellier II, 2009.

[9] P.-H. Cournède, M.Z. Kang, A. Mathieu, J.-F. Barczi, H.P. Yan, B.G. Hu, P. de Reffye, *Structural factorization of plants to compute their functional and architectural growth*. Simulation, 82 (2006), No. 7, 427–438.

[10] P.-H. Cournède, A. Mathieu, F. Houllier, D. Barthélémy, P. de Reffye. *Computing competition for light in the GreenLab model of plant growth: a contribution to the study of the effects of density on resource acquisition and architectural development*. Annals of Botany, 101 (2008), No. 8, 1207–1219.

[11] H. Cramer. Mathematical methods of statistics. Princeton University Press, Princeton, 1946.

[12] P. de Reffye, F. Blaise, S. Chemouny, T. Fourcaud, F. Houllier. *Calibration of hydraulic growth model on the architecture of cotton plants*. Agronomie, 19 (1999), 265–280.

[13] P. de Reffye, E. Elguero, E. Costes. *Growth units construction in trees: a stochastic approach*. Acta Biotheoretica, 39 (1991), 325–342.

[14] P. de Reffye, E. Heuvelink, D. Barthélémy, P.-H. Cournède. *Plant growth models*. Ecological Models, Vol. 4 of Encyclopedia of Ecology (5 volumes) (S.E. Jorgensen and B. Fath, eds.), Elsevier, Oxford, 2008, 2824–2837.

[15] P. de Reffye, S. Lemaire, N. Srivastava, F. Maupas, P.-H. Cournède. *Modeling inter-individual variability in sugar beet populations*. 3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China (B.G. Li, M. Jaeger, Y. Guo, eds.), IEEE, November 9-12 2009.

[16] A.P. Dempster, N.M. Laird, D.B. Rubin. *Maximum Likelihood from incomplete data via the EM algorithm*. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 39 (1977), 1–38.

[17] A. Doucet, A.M. Johansen. *A tutorial on particle filtering and smoothing: fifteen years later*. Tech. report, Department of Statistics, University of British Columbia, 2008.

[18] J.-L. Drouet, L. Pagès. *GRAAL: a model of GRowth Architecture and carbon ALlocation during the vegetative phase of the whole maize plant, model description and parameterisation*. Ecological Modelling, 165 (2003), 147–173.

[19] C. Eschenbach. *Emergent properties modelled with the functional structural tree growth model almis: Computer experiments on resource gain and use*. Ecological Modelling, 186 (2005), No. 4, 470 – 488.

[20] J.F. Farrar. *Sink strength: What is it and how do we measure it? A summary*. Plant, Cell and Environment, 16 (1993), No. 9, 1045–1046.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

[21] T. Fourcaud, X.P. Zhang, A. Stokes, H. Lambers, C. Körner. *Plant growth modelling and applications: The increasing importance of plant architecture in growth models*. Annals of Botany, 101 (2008), No. 8, 1053–1063.

[22] C. Gaucherel, F. Campillo, L. Misson, J. Guiot, J.-J. Boreux. *Parameterization of a process-based tree-growth model: Comparison of optimization, MCMC and particle filtering algorithms*. Environmental Modelling and Software, 23 (2008), No. 10-11, 1280 – 1288.

[23] C. Godin, H. Sinoquet. *Functional-strucutural plant modelling*. New Phytologist, 166 (2005), 705–708.

[24] G.C. Goodwin, R.L. Payne. Dynamic system identification: Experiment design and data analysis. Academic Press, New York, 1977.

[25] Y. Guédon, D. Barthélémy, Y. Caraglio, E. Costes. *Pattern analysis in branching and axillary flowering sequences*. Journal of Theoretical Biology, 212 (2001), 481–520.

[26] H. Guo, V. Letort, L. Hong, T. Fourcaud, P.-H. Cournède, Y. Lu, P. de Reffye. *Adaptation of the Greenlab model for analyzing sink-source relationships in Chinese pine saplings*. Plant growth Modeling, simulation, visualization and their Applications (PMA06). (T. Fourcaud and XP. Zhang, eds.), IEEE Computer Society, Los Alamitos, California, 2007.

[27] Y. Guo, Y.T. Ma, Z.G. Zhan, B.G. Li, M. Dingkuhn, D. Luquet, P. de Reffye. *Parameter optimization and field validation of the functional-structural model Greenlab for Maize*. Annals of Botany, 97 (2006), 217–230.

[28] F. Hallé, R.A.A. Oldeman. Essai sur l'architecture et la dynamique de croissance des arbres tropicaux. Masson, Paris, 1970.

[29] G. Hammer, M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. Van Eeuwijk, S. Chapman, D. Podlich. *Models for navigating biological complexity in breeding improved crop plants.* Trends in Plant Science, 11 (2006), No. 12, 587–593.

[30] E. Heuvelink. *Evaluation of a dynamic simulation model for tomato crop growth and development*. Annals of Botany, 83 (1999), 413–422.

[31] C.A. Jones, J.R. Kiniry. *Ceres - Maize : A simulation model of Maize growth and development*. Texas A&M University Press, 1986.

[32] S. Julier, J. Uhlmann, H.F. Durrant-Whyte. *A new method for the nonlinear transformation of means and covariances in filters and estimators*. IEEE Transactions on Automatic Control, 45 (2000), No. 3, 477–482.

[33] M.Z. Kang, P.-H. Cournède, P. de Reffye, D. Auclair, B.G. Hu. *Analytical study of a stochastic plant growth model: application to the Greenlab model*. Mathematics and Computers in Simulation, 78 (2008), No. 1, 57–75.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

[34] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. *Optimization by Simulated Annealing*. Science, 220 (1983), No. 4598, 671–680.

[35] A. Lacointe. *Carbon allocation among tree organs : A review of basic processes and representation in functional-structural models*. Annals of Forest Sciences, 57 (2000), 521–533.

[36] S. Lemaire, F. Maupas, P.-H. Cournède, P. de Reffye. *A morphogenetic crop model for sugarbeet (Beta Vulgaris l.)*. Crop Modeling and Decision Support, (W. Cao, J. White, E. Wang, eds.), Springer, 2009, 116–129.

[37] V. Letort. Multi-scale analysis of source-sink relationships in plant growth models for parameter identification. Case of the Greenlab model. Ph.D. thesis, Ecole Centrale Paris, 2008.

[38] V. Letort, P.-H. Cournède, A. Mathieu, P. de Reffye, T. Constant. *Parametric identification of a functional-structural tree growth model and application to beech trees (Fagus Sylvatica)*. Functional Plant Biology, 35 (2008), No. 10, 1243–1254.

[39] V. Letort, P. Mahe, P.-H. Cournède, P. de Reffye, B. Courtois. *Quantitative genetics and functional-structural plant growth models: Simulation of quantitative trait loci detection for model parameters and application to potential yield optimization*. Annals of Botany, 101 (2008), No. 8, 951–963.

[40] L. Ljung. *Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems*. IEEE Trans. Autom. Control, 20 (1979), No. 5, 643–652.

[41] L. Ljung. System identification: Theory for the user. Prentice-Hall, New Jersey, 1999.

[42] L. Ljung, T. Söderström. Theory and practice of recursive identification. MIT Press, Cambridge, MA, 1983.

[43] C. Loi, P.-H. Cournède. *Generating functions of stochastic L-systems and application to models of plant development*. Discrete Mathematics and Theoretical Computer Science Proceedings, AI (2008), 325–338.

[44] C. Loi, P.-H. Cournède, J. Françon. *A symbolic method to analyse patterns in plant structure*. Journal of Computer Science and Technology, In Press (2011).

[45] D. Luquet, M. Dingkuhn, H. Kim, L. Tambour, A. Clément-Vidal. *Ecomeristem, a model of morphogenesis and competition among sinks in rice. 1. concept, validation and sensitivity analysis*. Functional Plant Biology, 33 (2007), No. 4, 309–323.

[46] Y.T. Ma, M.P. Wen, B.G. Li, P.-H. Cournède, P. de Reffye. *Calibration of Greenlab model for Maize with sparse experimental data*. The Second International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA06) (T. Fourcaud and X.P. Zhang, eds.), IEEE, Los Alamitos, California, 2007.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

[47] D. Makowski, J. Hillier, D. Wallach, B. Andrieu, M.-H. Jeuffroy. *Parameter estimation for crop models*. Working with Dynamic Crop Models, (D. Wallach, D. Makowski, J.W. Jones, eds.), Elsevier, 2006, 55–100.

[48] A. Mathieu, P.-H. Cournède, V. Letort, D. Barthélémy, P. de Reffye. *A dynamic model of plant growth with interactions between development and functional mechanisms to study plant structural plasticity related to trophic competition*. Annals of Botany, 103 (2009), 1173–1186.

[49] A. Mathieu, B.G. Zhang, E. Heuvelink, S.J. Liu, P.-H. Cournède, P. de Reffye. *Calibration of fruit cyclic patterns in cucumber plants as a function of source-sink ratio with the Greenlab model*. Proceedings of the 5th international workshop on FSPM (P. Prusinkiewicz, J. Hanan, eds.), November 2007.

[50] E. Nicolini, B. Chanson, F. Bonne. *Stem growth and epicormic branch formation in understorey beech trees (Fagus Sylvatica l.)*. Annals of Botany, 87 (2001), 737–750.

[51] J. Olsson, O. Cappé, R. Douc, E. Moulines. Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models. Bernoulli, 14 (2008), No. 1, 155–179.

[52] B. Pallas, A. Christophe, P.-H. Cournède, J. Lecoeur. *Using a mathematical model to evaluate the trophic and non-trophic determinants of axis development in Grapevine (Vitis Vinifera l.)*. Functional Plant Biology, 36 (2009), No. 2, 156–170.

[53] P. Prusinkiewicz. *Modeling plant growth and development*. Current Opinion in Plant Biology, 7 (2004), No. 1, 79–84.

[54] M. Quach, N. Brunel, F. d'Alché Buc. *Estimating parameters and hidden variables in non-linear state-space models based on ODEs for biological networks inference*. Bioinformatics, 23 (2007), No. 23, 3209–3216.

[55] L.R. Rabiner. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proc. IEEE, 77 (1989), 257–284.

[56] O. Schabenberger, F. Pierce. Contemporary statistical models for the plant and soil sciences. Addison-Wesley, 2001.

[57] R. Sievänen, E. Nikinmaa, P. Nygren, H. Ozier-Lafontaine, J. Perttunen, H. Hakula. *Components of a functional-structural tree model*. Annals of Forest Sciences, 57 (2000), 399–412.

[58] W. Taylor. *Small sample properties of a class of two-stage Aitken estimator*. Econometrica, 45 (1977), No. 2, 497–508.

[59] B Thiébaut, S. Puech. Développement du hêtre commun. Morphologie et architecture de l'arbre. Revue Forestière Française, 56 (1984), No. 1, 45–58.

[60] E. Walter, L. Pronzato. Identification de modèles paramétriques. Masson, Paris, 2006.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press

[61] F. Wang, M.Z. Kang, Q. Lu, H. Han, V. Letort, Y. Guo, P. de Reffye, B. Li. *Simulation of the structure and function of young Mongolian scots pine trees using the stochastic Greenlab model*. Annals of Botany, In press (2010).

[62] J. Warren-Wilson. *Ecological data on dry matter production by plants and plant communities*. The collection and processing of field data (E.F. Bradley, O.T. Denmead, eds.), Interscience Publishers, New York, 1967, 77–123.

[63] P. Wernecke, J. Müller, T. Dornbusch, A. Wernecke, W. Diepenbrock. *The virtual crop modelling system VICA specified for Barley*. Functional-structural plant modelling in crop production, (J. Vos, L.F.M. Marcelis, P.H.B. de Visser, P.C. Struik, J.B. Evers, eds.), Springer, 2007, 58–69.

[64] Q. Wu, J. Bertheloot, A. Mathieu, B. Andrieu, P.-H. Cournède. *Assessment of non-linearity in functional-structural plant models*. 6th international workshop on Functional-Structural Plant Models (FSPM10), Davis, USA (T. De Jong, J. Vos, A. Escobar, eds.), November 9-12 2010.

[65] Q. Wu, P.-H. Cournède. *Sensitivity analysis of Greenlab model for Maize*. 3rd international symposium on Plant Growth and Applications(PMA09), Beijing, China (B.G. Li, M. Jaeger, Y. Guo, eds.), IEEE, November 9-12 2009.

[66] H.P. Yan, M.Z. Kang, P. de Reffye, M. Dingkuhn. *A dynamic, architectural plant model simulating resource-dependent growth*. Annals of Botany, 93 (2004), 591–602.

[67] Z.G. Zhan, P. de Reffye, F. Houllier, B.G. Hu. *Fitting a structural-functional model with plant architectural data*. Plant Growth Models and Applications (B.G. Hu, M. Jaeger, eds.), Tsinghua University Press and Springer (Beijing, China), 2003, 236–249.

[68] W. Zucchini, I.L. MacDonald. Hidden Markov models for time series - An introduction using R. Chapman and Hall, 2009.

https://doi.org/10.1051/mmnp/20116205 Published online by Cambridge University Press