

Fitting a Functional-Structural growth model with plant architectural data

Z. G. Zhan, Philippe de Reffye, François Houllier, B.G. Hu

▶ **To cite this version:**

Z. G. Zhan, Philippe de Reffye, François Houllier, B.G. Hu. Fitting a Functional-Structural growth model with plant architectural data. International Symposium on Plant Growth Modeling, Simulation, Visualization and their Applications - PMA'03, Oct 2003, Beijing / China, China. pp.108-117. inria-00122502

HAL Id: inria-00122502

https://inria.hal.science/inria-00122502

Submitted on 12 Jan 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fitting a Functional-Structural Growth Model with Plant Architectural Data

Z.-G. Zhan[1, 2*], P. De Reffye[1,2, 3], F. Houllier[3], B.-G. Hu[1]

[1]*LIAMA, Institute of Automation, Chinese Academy of Sciences, China*
[2]*INRIA Rocquencourt, France*
[3]*AMAP, A Joint Research Unit on Botany and Plant Architecture Bioinformatics, France*

Abstract

GreenLab is a recurrent discrete-time functional-structural model of plant growth and architecture. A method is presented estimating its parameters: the model is fitted to plant morphological and architectural data observed at one point of time. Since GreenLab output variables (number, size and fresh mass of organs) implicitly and nonlinearly depend on the model parameters, the fitting problem is solved by minimizing a generalized least-squares criterion and by implementing an iterative procedure. Fitting is satisfactorily performed on unbranched plants (cotton, maize, sunflower) using real data. The method is extended to more complex plants (i.e. with branches): a preliminary test on a virtual tree shows that the fitting algorithm also applies to such structured plants.

Keywords: nonlinear fitting, least squares method, modeling, plant architecture, plant growth

1 Introduction

In the first process-based plant growth models, plant architecture was oversimplified [1, 2]: plant organs were aggregated into broad compartments (e.g. leaf, branch, stem or root compartments), and their geometrical and topological position was not taken into account. The role of interactions among organs and the need of models in agronomy have lead to a stronger coupling of ecophysiological processes and plant structure, which requires a more detailed representation of plant architecture [1, 3, 12, 13].

The GreenLab model, developed by LIAMA, CIRAD and INRIA, and derived from AMAP models [2, 4, 5], is based upon several simple and general principles that account for morphogenetic and ecophysiological processes at an integrated time step: the production of new organs is driven by the accumulated temperatures; the water transpiration of the leaves depends on the hydraulic architecture of the plant, while the biomass production and the growth of organs depend on carbon assimilation and average water use efficiency, carbon allocation (in fact, biomass partitioning) and organ geometrical allometries.

*Corresponding author, e-mail: zgzhan@liama.ia.ac.cn

The aim of this paper is to show that, and how, GreenLab's parameters can be estimated by fitting the model to classical architectural and morphological data measured at only one point of time. We briefly describe the underlying functional-structural growth model. Since its functional parameters cannot be directly measured, we propose an extended least squares method to estimate these "hidden" parameters from field measurements. We illustrate this method on some simple unbranched or pruned plants: cotton, maize and sunflower. We then suggest that this method can be extended to more complex plants (i.e. branched trees): this extension is tested using simulated data, instead of measurements.

2 The GreenLab Model Description

2.1 Biological assumptions

A GreenLab virtual plant is assumed to be made of fresh matter, which contains about 80% of H_2O and 15% of carbohydrates (CO_2 + H_2O) [5]. The model only operates on the aerial part of the plant, whose mass is assumed to be proportional to that of the root system. The plant growth simulation is based on an architectural description of the plant in terms of organs (e.g. internodes, leaves, flowers, fruits, etc.) and plant morphogenesis is simulated as an automaton with respect to the species architectural model [6], while fresh matter production and the size of organs are computed according to global ecophysiological processes based on two agronomic laws (see below).

Morphogenesis and the sum of temperatures. Plants are supposed to grow in cycles, whose duration may range from several days (e.g. cotton's cycle varies from 3 to 7 days) to one year (e.g. for temperate tree species) [5]. The number of growth cycles undergone by a bud during a given time period is supposed to be proportional to the sum of the daily average temperature above a given threshold. For a given species and environment, this law determines the correspondence between real time and the number of growth cycles, thus allowing to furthers reason on growth cycles (i.e. thermally adjusted cycles). Thus, the speed of the morphogenetic automaton, the topological structure of the plant and the number of its organs depend on the cumulative effect of the physical environment.

Biomass production and the water use efficiency. Over a sufficient time lag (e.g. the duration of the culture), and subject to a given amount of incoming light, the amount of fresh matter accumulated by the plant is proportional to its cumulated water transpiration [5]. The proportionality factor is called the water use efficiency, and varies according to the biological type of the plant. Although water use efficiency may vary over a short period and for a particular leaf [12], we consider here its average value among all organs and over time. We further assume that there is no water stress and use this proportionality to derive the amount of matter produced from the computed water transpiration at the whole plant level.

2.2 Modeling fresh matter production

In order to compute the fresh matter production we need to calculate plant transpiration. Water is taken from the ground by the roots and flows through the axes into the leaves where it is transpired. The distribution of this flow is assumed to obey simple network rules, which are based on the topological arrangement of plant components and on their hydraulic resistance. It is assumed that the major hydraulic resistance of the plant lies in the leaves, and that the set of active leaves operate in parallel. As a result, during the ith growth cycle, the matter production only depends on the number of leaves $n(i)$ that are available for

photosynthesis. This number results from the previous development of the plant and depends on the architectural model of the species [6]. If each leaf receives the same amount of energy and undergoes the same microclimatic conditions, the biomass production at cycle i is:

$$Q_m(i) = \sum_{j=1}^{n(i)} \frac{E(i)}{r_1 / s(j) + r_2} \tag{1}$$

where $Q_m(i)$ is the biomass production during cycle i; $E(i)$ is the average potential of biomass production during cycle i ($E(i)$ depends on the water use efficiency and on physical variables such as temperature, light availability, soil water availability, etc.); $s(j)$ is the blade surface of the jth leaf; r_1 is the leaf blade resistance per unit area (note the inverse variation of resistance with area) and r_2 is the average resistance of the network constituted by the nerves and petiole of the leaf.

2.3 Modeling biomass allocation and organ expansion

During plant growth, all organs are considered as sinks that compete with each other for the produced biomass (or the photosynthetically assimilated carbon), according to their strength. It is assumed that the expansion rate of organs depends both on their type and thermal age (i.e. their age in growth cycles). Therefore, the sink strength for an organ denoted by o is defined as a function of its thermal age:

$$G_o(j) = P_o \cdot f_o(j) \tag{2}$$

where o=a, e, c, f, m refers to organ type, respectively to leaf, internode, layer, female organ (flower and fruit) and male organ; P_o is a constant coefficient of sink strength associated with organ type o; $f_o(j)$ is a function of the thermal age j of the organ, under a normalization constraint $\sum_{j=1}^{t_o} f_o(j) = 1$; t_o is the expansion duration in cycles for organ type o.

We assume that function f_o has the following form:

$$\begin{cases} f_o(j) = \begin{cases} g_o(j)/M_o & (1 \le j \le t_o) \\ 0 & (j > t_o) \end{cases} \\ g_o(j) = (j-0.5)^{\alpha_o - 1} \cdot (t_o - j + 0.5)^{\beta_o - 1} \\ M_o = \sum_{k=1}^{t_o} g_o(k) \end{cases} \tag{3}$$

where the parameters α_o and β_o vary with the organ type. We define $C_o = \alpha_o + \beta_o$ and $B_o = \alpha_o / C_o$. The parameter C_o is fixed to a certain value and B_o is to be estimated.

Let $N_o(i,j)$ be the number of organs of type o whose thermal age is j at cycle i. At the beginning of cycle i, the biomass increment of an individual organ of type o, defined as $\Delta q_o(i,j)$, is computed as

$$\Delta q_o(i,j) = \frac{G_o(j)}{\sum_{o=a,e} \sum_{j=1}^{i} N_o(i,j) \cdot G_o(j) + \sum_{o=c,f,m} \sum_{j=1}^{i-1} N_o(i,j) \cdot G_o(j)} \cdot Q_m(i-1) \tag{4}$$

where $Q_m(i-1)$ is the biomass available (i.e. produced) at the beginning of cycle i for all sinks: internodes, leaves, etc.

It is assumed that the formation of a new layer along the existing internodes as well as the

growth of female and male organs happen at the end of the cycle. Therefore, at the same node, the reproductive organs are always one cycle younger than the leaves (see Eq.(4)).

Let $q_o(i, j)$ be the total biomass of an individual organ of type o with thermal age j at cycle i. It is obtained by accumulating all the biomass increments from its birth to current cycle:

$$q_o(i,j) = \sum_{k=1}^{j} \Delta q_o(i-j+1,\ k) \tag{5}$$

Note that the root system is ignored if no data are available. GreenLab model can process it as an indefinitely growing organ whose biomass is controlled by a variable sink. Moreover we assume that the memory of the reserves is limited to only one growth cycle.

2.4 Computation of organ geometry

The size and geometry of organs are calculated from fresh masses according to allometric rules. Although its shape may be arbitrary, the blade area has a good linear relationship with its mass [5], [7]. The ratio "blade mass: whole leaf mass" is quite stable for given plant [5]. Petioles and internodes are considered as cylinders. The length and section of an internode are obtained from its volume:

$$\begin{cases} l_e = b_e \cdot V_e^{a_e} \\ \sigma_e = V_e \cdot l_e^{-1} \end{cases} \tag{6}$$

where V_e is the volume of the internode, l_e the length, and σ_e the section; b_e, a_e are two allometric parameters. The organ volume has the same value as its mass if the organ density is 1.

Similar formula are used for calculating the geometry of the petiole, whose allometric parameters are b_p and a_p. Secondary growth is modeled as the addition of 'layers' along the axes. It is further assumed that these layers are uniformly distributed along the axes. For example, the addition of one layer of biomass q_c, covering a stem whose length is h, generates an increment in the radius of the internode from R_{i-1} to R_i:

$$R_i = \sqrt{R_{i-1}^2 + \frac{q_c}{\pi \cdot h}} \tag{7}$$

2.5 The simulation algorithm

The numbers of organs (i.e. the number of functioning leaves, $n(i)$, and the number of organs of type o, $N_o(i, j)$) are critical variables in Equations (1) and (4). All these numbers are associated with plant topology and vary with time according to the architecture of the species [6]. A 2-scale automaton model [8] has been integrated into GreenLab to simulate the development of the plant. This automaton relies on the decomposition of plants into four types of basic elements: metamers, which are made of an elementary portion of an axis which bears lateral organs (i.e. leaves, flowers, fruits and lateral buds); growth units, which are constituted of the succession of metamers; axes which are made of the succession of growth units; structures which are formed by the assemblage of axes that are branched on each other (see Fig.3 for a simple example).

In this automaton, the axes are categorized according to two concepts of age, which should be clearly distinguished. The "chronological age" (or "thermal age") of an axis is defined as the number of growth cycles since its appearance, while its "physiological age" [14] accounts for its position within plant architecture (e.g. the physiological age can be linked to the morphological and vigor gradients, for example to branch order in some tree

species) and is the same for all growth units which form an axis. As a consequence, the chronological age of an axis varies with time, whereas its physiological age remains constant (i.e. along an axis, all growth units and metamers have the same physiological age, but the axes that are branched on this axis may have a different physiological age). The morphological and functional parameters of GreenLab are supposed to be identical for all organs that have the same physiological age. Note that these parameters are therefore independent of the chronological age.

In this model, the feedback between plant architecture, plant functioning and plant growth is automatically built-in at the level of growth cycle. The plant growth starts from the seed, which has a biomass $Q(0)$ at the beginning of the first cycle; the seed is used for constructing the first metamer; then the plant begins to grow on its own by interacting with its environment: biomass is produced, new organs are created and existing organs expand; and so on. The simulation procedure is described below, for growth cycle i:

- at the beginning of cycle, the plant creates new organs according to its architectural model [6], and expands its formerly created organs;
- during the cycle, the plant keeps its architecture and the functioning leaves fabricate fresh matter by photosynthesis;
- at the end of the cycle, secondary growth occurs and new layers (if any) are added to internodes. The reproductive organs can also appear at that time.

3 The Method for Estimating Parameters

Before using the model in simulation, prediction or control studies, it is necessary to estimate its parameters for a given species and in a given ecological context. These parameters are classified into two categories: those which can be simply estimated from direct architectural and morphological plant measurements, and the "hidden" parameters which cannot be directly estimated because they are deeply implied in the recursive simulation process described above. The parameters of the architectural automaton (e.g. the number of organs per metamer) and the allometric parameters of organ geometry (e.g. blade thickness, scaling coefficients for petiole and that for internode) belong to the first category: they can be obtained by classical plant architecture analysis methods [15] or by simple data analysis of morphological organ measurements (e.g. see [5, 7, 8]). In contrast, the functional parameters (i.e. the parameters associated to sink strength and to hydraulic resistance) cannot be obtained directly from measured architectural and morphological data because these parameters are used in a recursive growth algorithm, as seen in Equations (1)-(5). A special algorithm is therefore required to estimate these hidden functional parameters.

3.1 Measurements of plant architecture and morphology

In this paper, we consider the estimation of the hidden parameters, in the particular case where architectural and morphological information are only available at one point of time. Since the measurement of a plant can be very time-consuming, there is a need to design a sampling strategy that avoids the complete measurement of the plant. Such a strategy is especially required for large tree species that are composed of many axes, which form a branched topological structure, have different morphological features (e.g. these branches may be more or less vigorous according to their position in the tree), are more or less old (i.e. all branches do not appear at the same time), and often exhibit stochastic variability.

This sampling strategy is based upon the fact that various parts of the plants are similar to each other, and more precisely upon the categorization of plant axes according to their "chronological age" and "physiological age". It consists in selecting several categories of axes and in measuring some architectural and morphological variables on some axes within each selected category. The nature and number of selected axis categories and of measured axes within these categories depend on plant size and complexity. For example, a species that belongs to Corner model has only one axis (i.e. the main stem) [6], so that this axis can be exhaustively measured. But plant structure is often much more complex than a single stem bearing several leaves, flowers and fruits. For such complex plants, the sampling strategy is crucial to reduce the amount of measurements. If plant organogenesis is well known, the sample needed to fit the model can be small because the same substructures can be repeated many times inside the plant architecture [11] and because the hidden model parameters only depend on the physiological age (they do not depend on the chronological age).

As a result of model structure and of the sampling strategy, the morphological and architectural data (including length, diameter, area, and fresh mass of each sampled organ) can be organized into a compound data structure:

$$\begin{cases} \text{Plant data} = \Sigma\{A_i\} \\ \text{Axis data } A_i = \{y_{v,u,i}\} \end{cases} \tag{8}$$

where A_i is the data set for the ith axis; $y_{v,u,i}$ is the elementary data where the index 'v' identifies the category of the variable (for example, 0 for internode diameter, 1 for internode length, 2 for internode mass, 3 for petiole diameter, etc.), and 'u' is the growth unit rank along the axis counted from the plant stump upwards. The number of variable categories is not necessarily the same for all the axes. Finally a vector Y made up of n measurement values is formed by arranging the data of m measured axes in one column:

$$\mathbf{Y} = \left[\mathbf{Y}^{\mathrm{T}}(A_1) \quad \cdots \quad \mathbf{Y}^{\mathrm{T}}(A_1) \quad \cdots \quad \mathbf{Y}^{\mathrm{T}}(A_m)\right]^{\mathrm{T}}_{n\times 1} \tag{9}$$

where $\mathbf{Y}(A_i) = (y_{v,u,i})^{\mathrm{T}}_{(Vi\times Ui)\times 1}$, $n = \sum_{i=1}^{m} V_i U_i$, V_i and U_i are respectively the number of variable categories and the number of growth units for the ith axis.

3.2 Application of the generalized least squares method

As a recurrent discrete model, GreenLab is not a linear model: its outputs (e.g. organ size and mass) cannot be expressed as a linear combination of the model parameters. However, the $\mathbf{Y}$ vector can be expressed as an implicit function of these parameters:

$$\mathbf{Y} = \mathbf{F}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{10}$$

where $\mathbf{F}(\boldsymbol{\theta}) = \left[f_1(\boldsymbol{\theta}) \quad \cdots \quad f_n(\boldsymbol{\theta})\right]^{\mathrm{T}}$ is a $n\times 1$ vector function with respect to $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_p]^{\mathrm{T}}$, a vector made of p unknown parameters; $\boldsymbol{\varepsilon}$ is a $n\times 1$ vector of zero-centered random variables.

Let $\boldsymbol{\Sigma}$ be the variance matrix of $\boldsymbol{\varepsilon}$, $\boldsymbol{\Sigma}$ is a $(n \times n)$ matrix. If the random errors are independent from each other, $\boldsymbol{\Sigma}$ is a diagonal matrix.

In order to get an estimate of $\boldsymbol{\theta}$, we use the weighted least squares criterion:

$$\mathbf{J}_{\Omega}(\boldsymbol{\theta}) = (\mathbf{Y} - \mathbf{F}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{\Omega} (\mathbf{Y} - \mathbf{F}(\boldsymbol{\theta})) \tag{11}$$

where $\boldsymbol{\Omega}$ is a $n\times n$ symmetric, positive matrix which weights the various observations.

Let $\mathbf{X}$ be the $n \times p$ matrix made of the partial derivatives of the model with respect to the unknown parameters:

$$\mathbf{X} = \left(\frac{\partial f_v(\boldsymbol{\theta})}{\partial \theta_k} \right)_{n \times p} \tag{12}$$

Since the model has no explicit analytical expression with respect to θ, these derivatives have to be numerically computed in Equation (13), which results in an approximation:

$$\frac{\partial f_v(\boldsymbol{\theta})}{\partial \theta_k} \approx \frac{\partial f_v(\boldsymbol{\theta} + h_k \mathbf{e}_k) - f_v(\boldsymbol{\theta})}{h_k} \tag{13}$$

where $\mathbf{e}_k$ is the kth column vector of the p×p identity matrix, and h_k is a finite-difference increment. A suggested choice for h_k [9] is:

$$h_k = \sqrt{\eta} \cdot |\theta_k| \tag{14}$$

where η is the machine precision of the computer.

Because the model is nonlinear with respect to $\boldsymbol{\theta}$, the minimization of $\mathbf{J}_{\Omega}(\boldsymbol{\theta})$ can only be performed through an iterative procedure. A primary guess estimate of $\boldsymbol{\theta}$, $\boldsymbol{\theta}^{(1)}$, is required to start the iterative process. Let consider the iterative procedure at stage q: near the estimated value $\boldsymbol{\theta}^{(q)}$ of $\boldsymbol{\theta}$, the model can be approximated by a linear model:

$$\begin{aligned} \mathbf{Y} &= \mathbf{F}(\boldsymbol{\theta}^{(q)}) + \mathbf{X}^{(q)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(q)}) + \boldsymbol{\varepsilon} \\ \mathbf{Z}^{(q)} &= \mathbf{Y} - \mathbf{F}(\boldsymbol{\theta}^{(q)}) = \mathbf{X}^{(q)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(q)}) + \boldsymbol{\varepsilon} \\ \mathbf{Z}^{(q)} &= \mathbf{X}^{(q)}\boldsymbol{\beta}^{(q)} + \boldsymbol{\varepsilon} \end{aligned} \tag{15}$$

where $\mathbf{Z}^{(q)}$ is the vector of deviations between the observed variables and their predicted values at step q; $\mathbf{X}^{(q)}$ is the sensitivity matrix computed in the vicinity of $\boldsymbol{\theta}^{(q)}$; $\boldsymbol{\beta}^{(q)} = \boldsymbol{\theta} - \boldsymbol{\theta}^{(q)}$ is the vector which measures the distance between the true parameters and their estimates at step q.

If the variance matrix $\boldsymbol{\Sigma}$ is a priori known or already well estimated, we use $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ as the weighting matrix. This has two advantages: it scales the variables to the same level, regardless of their units and, more importantly, it yields the best linear unbiased estimates of $\boldsymbol{\theta}$. In this case, we first estimate $\boldsymbol{\beta}^{(q)}$ from the approximated linear model and thus get $\boldsymbol{\theta}^{(q+1)}$ [10], a new estimate of $\boldsymbol{\theta}$:

$$\begin{aligned} \mathbf{J}^{(q)}(\boldsymbol{\theta}) &= (\mathbf{Z}^{(q)} - \mathbf{X}^{(q)}\boldsymbol{\beta}^{(q)})' \boldsymbol{\Sigma}^{-1} (\mathbf{Z}^{(q)} - \mathbf{X}^{(q)}\boldsymbol{\beta}^{(q)}) \\ \boldsymbol{\beta}^{(q)} &= (\mathbf{X}^{(q)'} \boldsymbol{\Sigma}^{-1} \mathbf{X}^{(q)})^{-1} \mathbf{X}^{(q)'} \boldsymbol{\Sigma}^{-1} \mathbf{Z}^{(q)} \\ \boldsymbol{\theta}^{(q+1)} &= \boldsymbol{\beta}^{(q)} + \boldsymbol{\theta}^{(q)} \end{aligned} \tag{16}$$

One difficulty arises if the variance matrix $\boldsymbol{\Sigma}$ is not a priori known. In this case, we propose a 2-step fitting procedure: first, in order to get rid of the effects of units and of the relative magnitude of the variables, we scale the data by multiplying them by the inverse of the average of the observations which belong to the same variable category; then, we fit the scaled model predictions with scaled data.

4 Some Results

4.1 Fitting GreenLab for real simple plants

Agricultural experiments were carried out in the China Agriculture University (2000-2002) with large spacing plantations of sunflower, maize and "single-stem cotton" (i.e. lateral buds

were systematically pruned once they appeared) [17]. In order to check GreenLab consistency and to test the above described estimation procedure, the aim of these experiments was to generate plants with a simple architecture. Plant architectural data collection was performed as explained in Section 3.1. Since the condition of no competition between plants and of no hydraulic stress was maintained during plant growth, the input variable $E(i)$ (the average potential of biomass production) can be set to a constant value of, say 1.0, as reference, merely for calibration purpose.

The model was fitted to these experimental data using the software GreenLab_*Fit*, which was especially designed for the calibration of such functional-structural models and which incorporates a GreenLab growth engine. The results are shown in Table 1 for the estimated values of the hidden parameters, in Fig.1 for the comparison between the observed data and the predicted values which result from the fitting, and in Fig.2 for the 3D plant architecture simulation. Note that the least squares method fits simultaneously all data histograms (i.e. all observed categories of variables on all observed axes are fitted at the same time). The plant geometry is predicted using the estimates of the hidden parameters.

Table 1 Estimated hidden parameters

	P_a	P_e	P_c	P_f	P_m	B_a	B_e	B_f	B_m	r_1	r_2
Cotton	1.0	0.10	0.30	-	-	0.37	0.25	-	-	31	0.30
Sunflower	1.0	0.045	0.041	8.2×10^2	-	0.43	0.24	0.73	-	37	0.075
Maize	1.0	1.3	-	3.7×10^2	0.88	0.35	0.73	0.77	0.5	39	0

Note: '-' means that the corresponding parameter is not used for modeling this species.

Fitting GreenLab on single stemmed cotton plant. The 6 parameters P_e, P_c, B_a, B_e, r_1, r_2 estimated on the 26-cycle old cotton plant are shown in Table 1. Note that $P_a = 1.0$ is a fixed *priori*, as a reference value for P_e and P_c, and that $C_a = 10, C_e = 5$ are also fixed. $t_a = t_e = 16$ are set according to observations.

Fitting GreenLab on sunflower. The sunflower ends its topological structure with a flower; hence, beside P_e, P_c, B_a, B_e, r_1 and r_2, the parameters associated to this organ, P_f and, B_f, have also to be estimated. The parameters are obtained in a similar way as for cotton. Fitting is made on the 63-cycle old plant. Note the sink strength of layer for sunflower is set to be proportional to the number of internode it covers, i.e. $G_c = P_c \cdot N_i$ where N_i is the internode number at present, while for cotton it is a invariant value, i.e. $G_c = P_c$.

Fitting GreenLab on maize. Maize is a finite structured plant with male and female organs set in different positions along the main stem. The fitting is focused on fresh masses and blade area. For maize, r_2 is set to zero. Parameters P_e, P_f, P_m, B_a, B_e, B_f, B_m, r_1 are estimated by fitting the model on the 30-cycle old plant.

4.2 Experiment of fitting with simulated data of a complex plant

In this section, we illustrate the possibility to extend the fitting algorithm to more complex plant. For deterministic branched plants, the application of the sampling strategy described in Section 3.1 indicates that sampling and measuring the plant stem and one axis per each pair of chronological age and physiological age may theoretically be sufficient to estimate the hidden parameters of the model.

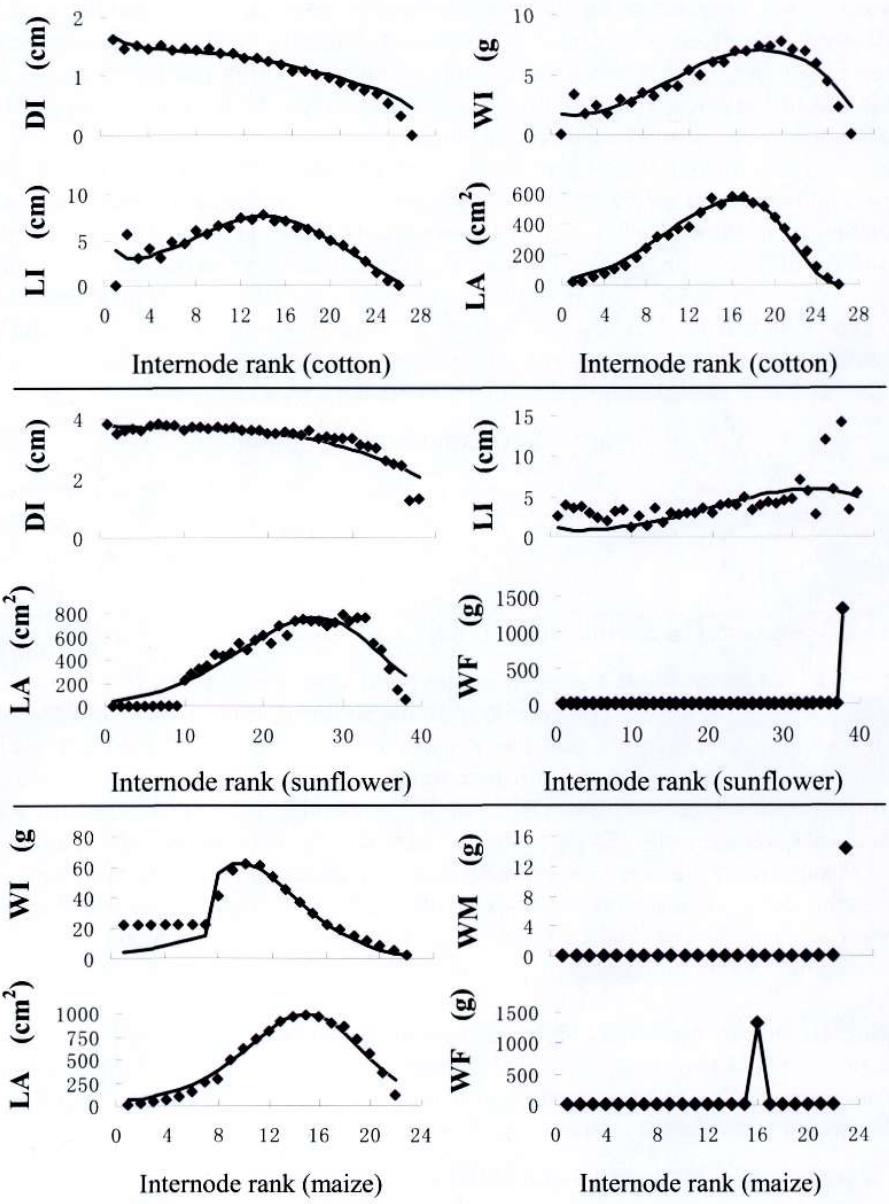


Fig.1 Fitting results on 3 single-stem plants: pruned cotton, sunflower and maize. Points marked with "◆" are measurement data, while simulation data are continuously linked lines (LA = Leaf area; LI = Length of internode; DI = Diameter of internode; WI = Weight of internode; WM = Weight of male organ; WF = Weight of female organ)

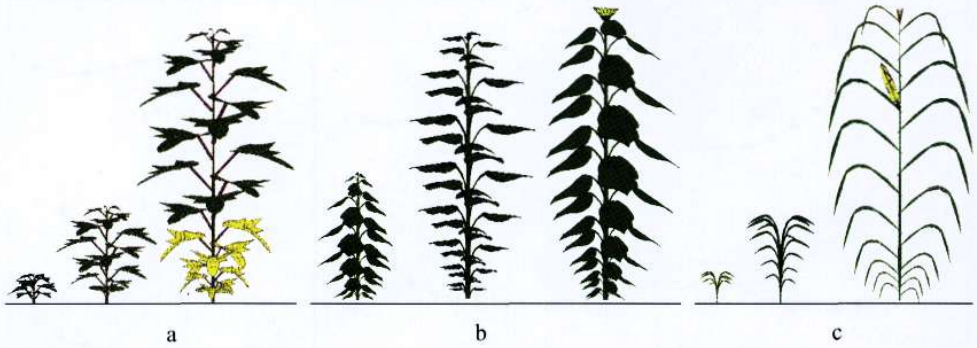


Fig.2 Reconstructed plant architectures. (a) pruned cotton at cycle 8, 16, 26; (b) sunflower at cycle 30, 42, 63; (c) maize at cycle 10, 15, 30

Here, consider a plant with a pure deterministic behavior, where the axes can have 3 different physiological ages (Fig.3). The leaves are of the same kind in all axes and their lifespan is equal to 3 growth cycles. Each growth unit of the main stem (i.e. the 'S_1 structure' whose physiological age is equal to 1) is made of 6 metamers, and each of the last 2 metamers bears a 'S_2 structure' (i.e. a main branch, or an axis whose physiological age is equal to 2). The growth units of the main branches are made of 4 metamers that bear a 'S_3 structure' (i.e. a twig, or a small branch whose physiological age is equal to 3). The growth units of the twigs are made of 2 metamers that bear 1 fruit.

We attempt to estimate 5 functional parameters (organ sinks) by fitting the model from only on 3 measured axes in the plant (i.e. the stem, 1 main branch and 1 twig). We used this simulated tree and considered the following situation: the simulated tree is 8 cycle-old. 3 axes are sampled: the main axis (i.e. the bearing axis of S_1 in Fig.3), a main branch of physiological age 2 and chronological age 7 (the bearing axis of S_2 in Fig.3); a twig of physiological age 3 and chronological age 6 (the bearing axis of S_3 in Fig.3). For these three axes we extract the following data: length and diameter of the internodes, area of the leaves and mass of the fruits. These data constitute the target data for the fitting algorithm (Fig.4).

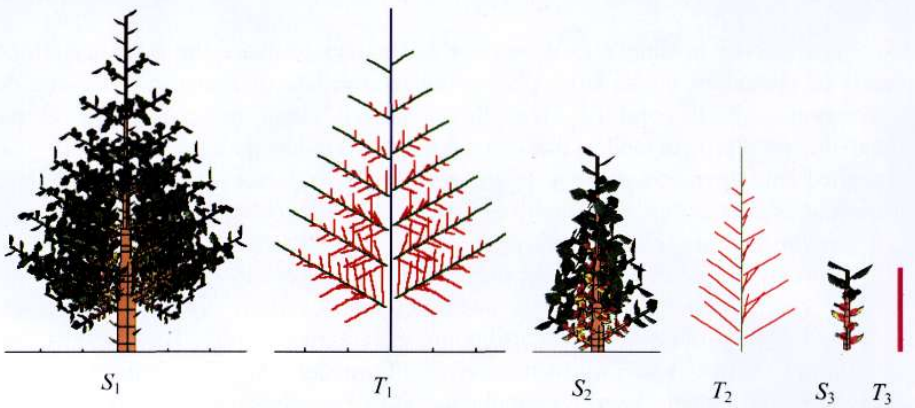


Fig.3 The architecture of a simulated deterministic tree. S_1 is the complete tree termed, its skeleton is T_1, the physiological age of the bearing axis is 1; S_2 is a main branch, it is a substructure of S_1, its skeleton is T_2, the physiological age of the bearing axis is 2; S_3 is a twig, it is a substructure of S_2, its skeleton is T_3, the physiological age of the bearing axis is 3

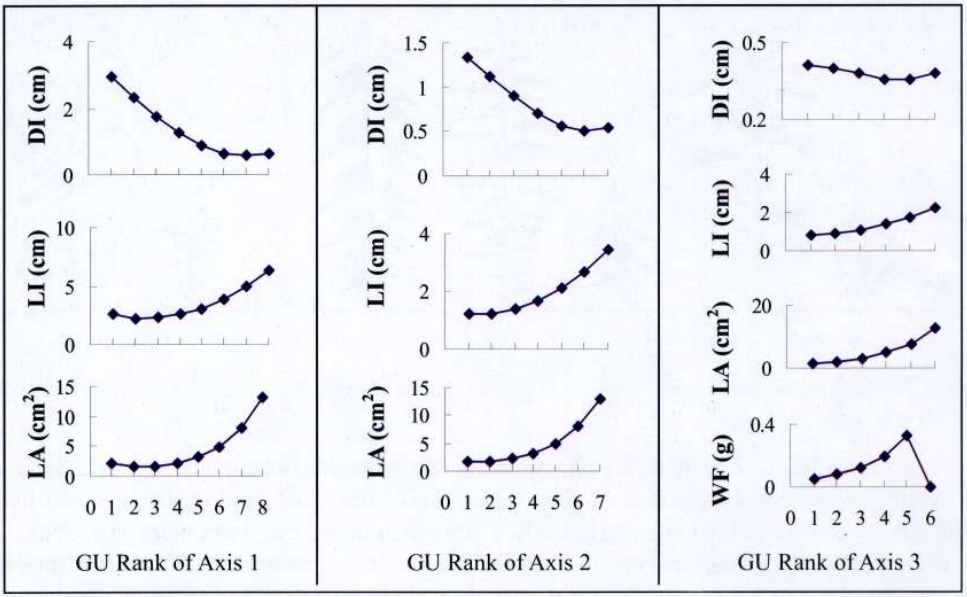


Fig.4 Target data of simulated tree for the fitting. GU = Growth Unit; LA = Leaf Area; LI = Length of Internode; DI = Diameter of Internode; WF = Weight of Fruit

Before starting the iterative procedure, the initial parameters are set to values that are different from those used for the simulation: the magnitude of the deviation is set to 20% of their true value. After a few iterations, the parameters converge to the theoretical values. This illustrates that, at least theoretically, a limited data set is sufficient for fitting the model. Before testing this algorithm on real plant, further methodological research should be carried out for testing it on complex theoretical plants, for example, using simulated data corrupted by noise.

5 Discussion and Conclusion

We have adapted the nonlinear least squares method to estimate the functional hidden parameters of GreenLab model from plant architectural data measured at only one date. From a methodological point of view, this approach could be generalized to other functional-structural growth models that contain unknown hidden parameters.

We applied this approach on real single-stem plants (i.e. unbranched plants): maize, sunflower and pruned cotton. Although these species are structurally different and exhibit different growth patterns (e.g. position and presence vs. absence of flowers), the fitting algorithm works well for each of them and the model appears to be enough generic and flexible for representing their growth throughout the various dynamic steps (plant development, biomass production and partitioning, organ growth, etc.). These results, added to preliminary studies with the same type of model [5, 7], indicate that this functional-structural model, based on simple and global ecophysiological principles, has a good potential for describing the major features of plant development for various species.

It should be noted that the applications presented in the paper deal only with isolated plants in optimal growth conditions and that, since the driving variable $E(\ )$ was set to a

constant value, say 1.0, the estimated values of parameters r_1 and r_2 have only a relative meaning. More experiments are required for the model validation, and also for the systematic analysis of residual errors for fitting results. Sticking with simple plants, further studies of this model and of the proposed fitting algorithm should indeed address three different issues.

First, the statistical properties of the estimates should be further studied analytically, by non-parametric techniques and/or by Monte-Carlo simulations. Indeed, the least squares method provides for also estimating, or at least approximating in the case of highly nonlinear models, the variance and covariance of estimation of the parameters [10]. It is however likely that these estimations will not be straightforward, because the recurrent nature of the model makes it highly nonlinear and generates a complex error structure (e.g. it is most likely that the observations are strongly autocorrelated along an axis). A careful statistical study might therefore require intensive Monte-Carlo simulations of various types of error structure as well as non-parametric techniques such as bootstrap.

Second, the fitting algorithm was applied to synchronic observations made at one point of time. It would therefore be important to check the dynamic consistency of the model. This could be done by at least two ways: either by verifying that the parameter estimates are constant when the model is fitted on data obtained at different ages; or by testing whether the predictions, retrospective or prospective, made with the model calibrated at one point of time match the observations at other dates. Earlier studies on wheat tend to indicate that GreenLab is indeed dynamically consistent and, therefore, that it could be applied for simulating plant growth over long time periods [7].

Third, the systematic fitting of various individuals of the same species, growing under either the same or different ecological conditions (i.e. water, light and temperature regimes), is required for testing the across-individuals and across-sites consistency of the model. On one hand, one would indeed expect that the estimates of the parameters of different plants grown under the same environment are fairly similar. On the other hand, if the model is robust, it is also expected that changes in the ecological conditions should not affect the functional parameters, provided appropriate scaling factors are introduced into the calibration procedure: changes in temperature should affect the speed of development (i.e. the relationship between normal time and thermal time) but not the estimates of the functional parameters associated to allocation pattern and to hydraulic resistances; on the other hand, water stresses and/or changes in light availability should affect parameter estimates because they are supposed to modify plant size through changes in E in Eq.(1).

If the model passes such tests, it may prove a highly valuable tool for comparing different plant genotypes of the same species, for analyzing the influence of environmental conditions and for controlling plant growth through light, water and/or temperature management or technical actions such as flower or fruit pruning [16]. On the contrary, if the model failed, it would be interesting to analyze which assumptions should be replaced or refined, for example, whether self-shading should be taken into account, or whether the allocation coefficients vary over time, or whether the assumption of a constant average water use efficiency does not hold on long time steps.

For the first time, the parameter estimation method was also extended to branched plants: the first test carried out on a virtual tree shows that this method can also be applied to such complex plants. A lot of work remains to be done with such plants: as for simple plants, the algorithm should be tested on real trees; the analysis of the statistical properties of the estimators remains should also be carried out along the lines proposed above for single-stem

species. In the case of trees, the design of efficient sampling strategy is also a major issue.

Acknowledgements

The authors thank the researchers of China Agriculture University: Yan Guo, Youhong Song, and Baogui Zhang for providing the experimental data. This work is supported in part by LIAMA, China NSFC(#60073007), and China 863 Program (#2002AA241221). The generous support from INRIA and CIRAD is gratefully acknowledged.

References

1 J.S. Hanan. "Virtual plants - integrating architectural and physiological plant models," *Environmental modeling and software*, vol. 12 (1), 1997, pp. 35–42.

2 Ph. (de) Reffye and F. Houllier. "Modeling plant growth and architecture: some recent advances and applications to agronomy and forestry", *Current science*, vol. 73 (11), 1997, pp. 984~992.

3 W. Kurth. "Morphological models of plant growth: possibilities and ecological relevance", *Ecological Modeling*, vol. 75/76, 1994, pp. 299-308.

4 Ph. (de) Reffye, Th. Fourcaud, F. Blaise, D. Barthélémy and F. Houllier. "An ecophysiological model for tree growth and tree architecture", *Silva Fennica*, vol. 31(3), 1997, pp. 297-311.

5 Ph. (de) Reffye, F. Blaise, S. Chemouny, S. Jaffuel, Th. Fourcaud and F. Houllier. "Calibration of hydraulic growth model on the architecture of cotton plants", *Agronomie*, vol.19, 1999, pp. 265-280.

6 F. Hallé, R.A.A. Oldeman, and P.B. Tomlinson. Tropical trees and forest: An architectural analysis, New York: Springer Verlag, 1978.

7 Z.G. Zhan, Y.M. Wang, Ph. (de) Reffye, B.B Wang and Y.W. Xiong. "Architectural Modeling of Wheat Growth and Validation Study", Presentation paper on ASAE 2000 Annual international meeting.

8 X. Zhao, Ph. (de) Reffye, F.L.Xiong, B.G. Hu and Z.G. Zhan. "Dual-scale automaton model for virtual plant development". *The Chinese Journal of Computers*, 2001, vol. 24(6), pp. 608-617

9 J.E. Dennis and R.B. Schnabel. Numerical methods for unconstrained optimization and nonlinear equations, New Jersey: Prentice-Hall, 1983.

10 S. Brandt. Statistical and computational methods in data analysis. North-Holland Publishing Company, 1976.

11 H.P. Yan, J.F. Barczi, Ph. (de) Reffye, B.G. Hu, M. Jaeger and J. Le Roux. "Fast algorithms of plant computation based on substructure instances". WSCG'2002, the 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2002.

12 H. Sinoquet and X. Le Roux. "Short-term interactions between tree foliage and the aerial environment: an overview of approaches available for tree structure-function models", *Annals of Forest Science*, 2000, vol. 57, pp. 477-496.

13 X. Le Roux, A. Lacointe, A. Escobar-Guttierez and S. Le Dizès. "Carbon-based models of individual tree growth: a critical appraisal", *Annals of Forest Science*, 2001, vol. 58, pp. 469-506.

14 D. Barthélémy, Y. Caraglio, E. Costes, " Architecture, gradients morphogénétiques et âge physiologique chez les végétaux", in *Modélisation et simulation de l'architecture des végétaux*, J. Bouchon, Ph. de Reffye and D. Barthélémy (Eds), Sciences Update, Editions INRA, Paris, 1997, pp. 89-136.

15 C. Godin, Y. Guédon, E. Costes, Y. Caraglio. "Measuring and analyzing plants with the AMAPmod software", in *Plants to ecosystems - Advances in Computational Life Sciences*, M.T. Michalewicz (Ed.), CSIRO Australia, 1997, pp. 53-84.

16 R. Nosenzo, Ph. (de) Reffye, F. Blaise and F. Le Dimet. "Principes de l'optimisation des modes de conduites culturales avec les modèles mathématiques de plantes", in *Modélisation des*

agro-écosystèmes et aide à la décision, E. Malézieux, G. Trébuil and M. Jaeger (Eds), CIRAD, Collection Repères, Montpellier, France, 2001, pp. 145-172.

17 Y. Guo, Ph. (de) Reffye, Y.H. Song, Z.G. Zhan, B.G. Li and M. Dingkuhn. "Modeling of biomass acquisition and partitioning in the architecture of sunflower", 2003' International symposium on plant growth modeling, simulation, visualization and their applications (PMA03).