SURGE 2009
Research Report
Modeling Heterogenity and Uncertainity Propagation in Plant Population Models for Parametric Identification of Sources of Inter-Individual Variability

Nitish Srivastava
Indian Institute of Technology, Kanpur, India
Mentor: Dr Paul-Henry Cournède
Applied Mathematics and Systems Lab, Ecole Centrale Paris, France

Abstract

We studied propagation of uncertainity in parameters of the GreenLab plant growth model and applied various techniques of estimation of Discrete Dynamic Systems to estimate the growth parameters required to model a plant population. Modeling heterogeneity in field crops is a key issue for a better characterization of field production. We choose a biologically plausible parameterized growth model for plants. The model is extended to plant populations. Several sources of individual variability in plant populations are identified, namely, initial conditions(seed mass, emergence delay), genetic variability(including phyllochron) and environment(incluing spacing and competition). A mathematical framework is introduced to integrate the various sources of variability in plant growth models. It is based on the method of Taylor Series Expansion, which allows the propagation of uncertainity in the dynamic system of growth and the computation of the approximate means and standard deviations of the model outputs. Parameter Estimation in models of a dynamic system generally involves adopting a probablistic framework for model equations by taking into account process and measurement errors. When system observations are regular, very efficient methods based on Kalman filtering have been devised, eg.extended Kalman filters in Ljung(1979) [2] or recursive least squares in Ljung and Söderström(1983) [3]. However some systems, such as living systems, do not allow regular data acquisition as measurements are too complex. For functional growth models, the system output is multivariate and generally corresponds to the masses of some elementary units of the plant. Measurements are thus destructive and can be done only once for a given plant. We attempt to give a method that relies on less data by including potential sources of strong variability across a plant population such as phyllochron and then using multifitting to fit the model at different stages of plant growth.

Contents

1 Brief Overview of the GreenLab model

GreenLab combines both organogenesis and functional growth in a mathematical dynamic system. A dual-scal automaton was initially used to describe GreenLab organogenesis. However, it was shown to be equivalent to a growth grammar(de Reffye(2003) [9]), and this formalism is mostly used from now on to describe the GreenLab model of plant development. In computational models, plants are generally represented as words in a formal language, more precisely a language based on a generative parallel rewriting grammar also called L-system(Lindenmayer(1968) [1], Smith(1984) [7], Prusinkiewicz et al. (1988) [5], Prusinkiewicz and Lindenmayer(1990) [4]). Even though the mathematical concepts underlying the GreenLab model were historically based on automaton theory, it has been shown that the model can be written in the framework of an L-System.

1.1 Plant Development and Growth Cycle

A metamer is a botanical entity chosen as the elementary scale to model plant architectural development in this study. It is composed of an internode bearing organs: axillary buds, leaves, flowers. The plant grows by successive shoots of several metamers produced by buds. The appearance of these shoots defines the architectural Growth Cycle. A Growth Unit is the set of metamers built by a bud during a growth cycle. We do not consider time scales that are smaller than the architectural growth cycle and we study the development of new growth units as a discrete process. The Chronological Age (CA) of a plant (or of an organ) is defined as the number of growth cycles it has existed for. Since metamers may bear axillary buds, plant architecture develops into a hierarchical branching system. Thus, the concept of Physiological Age (PA) was introduced to represent the different types of growth units and axes.The apical meristem or bud of an axis is thus characterized by the physiological age of the growth unit that it may produce and a metamer is characterized by its physiological age i (which is the physiological age of the growth unit that it belongs to) and that of the buds that it bears j.

1.2 Alphabet

In GreenLab, the alphabet $\mathcal{G}$ is given by the set of metamers $\mathcal{M}$ and buds $\mathcal{B}$ A metamer is defined with four indices and is denoted by $m_{pq}^{t}(n)$:

- its chronological age n
- its physiological age p,
- physiological age of its axillary buds q,$(q \geq p)$
- the chronological age of the plant t

A bud is defined by 3 indies and is denoted by $b_p^t(n)$:

- its physiological age p,
- the number of growth cycles n for which the bud's physiological age has been p - the ontogenic age of a bud.
- the chronological age of the plant t

The organogenesis alphabet is given by:

$$\mathcal{G} = \mathcal{M} \cup \mathcal{B} \tag{1.1}$$

with

$$\mathcal{M} = \{m_{pq}^{t}(n) \mid 1 \leq p \leq P, p \leq q \leq P, 1 \leq t \leq T, 1 \leq n \leq t\}, \tag{1.2}$$

and

$$\mathcal{B} = \{b_p^t(n) \mid 1 \leq p \leq P, 0 \leq t \leq T, 0 \leq n \leq t\}. \tag{1.3}$$

T denotes the maximum time for the simulation.

Definition(Set of words over an alhabet) The set of words over an alphabet $\mathcal{A}$ is defined as the monoid generated for the concatenation operator (.), seen as an internal, non-commutative operation, by $\mathcal{A} \cup \{1\}$, where 1 is the neutral element for the concatenation which corresponds to the empty word. It is denoted by $\mathcal{A}^*$

$\mathcal{G}^*$ will thus represent all the possible topological structures composed with buds and metamers.

1.3 L-Systems

Since plant development can be seen as the result of bud-functioning, generative grammars describing the production of buds have shown to be particularly adapted to mathematically model plant development. We recall below the definition of the simplest type of parallel generative grammars called 0L-System and first introduced by Lindenmayer(1968) [1].

Let $\mathcal{V} = \{v_1, v_2, \cdots, v_m\}$ denoe an alphabet and $\mathcal{V}^*$ the set of words over $\mathcal{V}$.

Definition(0L System) A 0L-system is a construct $G = \langle \mathcal{V}, w_a, P_r \rangle$where:

- $w_a \in \mathcal{V}^*$ is a word(called the axiom) which represents the structure initiating the growth.
- $P_r \subset (\mathcal{V} \cup \{1\} \times \mathcal{V}^*)$ is a finite set of productions and corresponds to the graph of a mapping from $\mathcal{V} \cup \{1\}$ into $\mathcal{V}^*$. We necessarily have $(1,1) \in P_r$. A production $(s, \chi) \in P_r$ is written as $s \to \chi$ and represents the evolution of the symbol s into χ. s is called a predecessor.

Directly inspired by this definiton a more general class of grammars was later introduced by Rozenberg and Salomaa(1980) [6], the F0L-systems.

Definition(F0L System) An F0L-system is a construct $G = \langle \mathcal{V}, W, P_r \rangle$where:

- $W \subset \mathcal{V}^*$ such that $Card(W)$ is finite, $W \neq \phi$ and $W \neq \{1\}$.
- $\forall\ w_a \in W, G[w_a] = \langle \mathcal{V}, w_a, P_r \rangle$ is a 0L-system.

W is called the set of axioms and $G[w_a]$ is called a component system of G.

The GreenLab organogenesis model can be modelled as a F0L-system as follows.

Definition(GreenLab Organogenesis)

GreenLab organogenesis is defined as an F0L system $\langle \mathcal{G}, \mathcal{B}, P_r \rangle$ with the following production rules P_r:

$\forall\ (t, n, p) \in [0, T] \times [0, \min(\tau(p), t)] \times [1, P]$:

$$b_p^t(n) \xrightarrow{if\ n<\tau(p)} \left(\prod_{q=p}^{P} \left(m_{pq}^{t+1}(1)\right)^{u_{pq}(t+1)} \left(b_p^{t+1}(0)\right)^{v_{pq}(t+1)} \right) b_p^{t+1}(n+1) \tag{1.4}$$

$$\xrightarrow{if\ n=\tau(p)} \left(\prod_{q=p}^{P} \left(m_{pq}^{t+1}(1)\right)^{u_{pq}(t+1)} \left(b_p^{t+1}(0)\right)^{v_{pq}(t+1)} \right) b_{\mu(p)}^{t+1}(0) \tag{1.5}$$

and $\forall\ (t, n, p, q) \in [1, T] \times [1, t] \times [1, P] \times [p, P]$:

$$m_{pq}^t \to m_{pq}^{t+1}(n+1)$$

with:

- $u_{pq}(t)$: number of phytomers m_{pq} in a growth unit of PA p, appearing at growth cycle t
- $v_{pq}(t)$: number of active axillary buds of PA q in a growth unit of PA p, appearing at growth cycle t
- $\tau(p)$: number of growth cycles after which a bud of PA p changes to PA $\mu(p)$. Vectors τ and μ characterize meristem differentiation.

2 Equations of Plant Growth

GreenLab describes the source-sink dynamicsduring plant growth. Sources correspond initially to the seed and then to the biomass production and reserve remobilization. Sinks are demands for biomass of all living organs. Since structural development is decribed in GreenLab at the level of organs, the computation of demand is coupled with organogenesis. For this reason a consistent time unit for architectural growth and photosynthetic production is defined in order to handle a constant structure. It allows the derivation of the discrete dynamic system of growth.

2.1 Fundamental GreenLab equation

$m^t_p(n)$ denotes at growth cycle t a metamer of physiological age p and chronological age n. It contains organs of type o ($o = b, p, r$ for blade, petiole or root respectively) whose masses are denoted by $q^t_{o,p}(n)$ at growth cycle t. Let $T_{o,p}$ denote the maximal life span of organs of type o and physiological age p. The biomass allocation equation is thus given for all $t \geq 0$ by:

$$q^{t+1}_{o,p}(n+1) = q^t_{o,p}(n) + \frac{p^t_{o,p}}{D^t}u^t \quad if\ 0 \leq n \leq \min(T_{o,p}-1, t) \tag{2.1}$$

$$q^t_{o,p}(n) = 0 \quad if\ n \geq min(T_{o,p}, t+1) \tag{2.2}$$

where:
$p^t_{o,p}(n)$ is the sink of an organ of type o in $m^t_p(n)$
D^t is the total demand of the plant at growth cycle t

$$D^t = \sum_{o,p} \sum_{n=0}^{T_{o,p}} N^t_{o,p}(n) p^t_{o,p}(n)$$

u^t is the biomass available for allocation.
It is to be noted that the allocation equation appears as the eulerain discretization of an ordinary diffrential equation of the form

$$\frac{d[q_{o,p}(n+1)](t)}{dt} = f(t, u(t)).$$

2.2 Sink function:p^t

Organ sink functions are chosen independent of the system state variables and identical for the organs of the same type and same physiological age. We suppose that the expansion of an organ of type o and physiological age p lasts from chronological age $T^i_{o,p}$ to chronological age $T^f_{o,p}$, and organ sink functions are modelled with flexible functions able to fit typical biological kinetics. Dicretized and normalized beta functions multiplied by the global sink value give good results. We choose, for $n \in [T^i_{o,p}, T^f_{o,p} - 1]$:

$$p^t_{o,p}(n) = P_{o,p} \left(\frac{n+0.5}{T^i_{o,p} - T^f_{o,p}} \right)^{\alpha_o} \left(1 - \frac{n+0.5}{T^i_{o,p} - T^f_{o,p}} \right)^{\beta_o} \frac{1}{max_{x \in [0,1]} x^{\alpha_o} (1-x)^{\beta_o}} \tag{2.3}$$

and $p^t_{o,p}(n) = 0$ otherwise. $P_{o,p}$ is the global organ sink value, α_o and β_o are the parameters of the beta function associated to the organ sink.

2.3 Available biomass:u^t

We denote by $\mathcal{A}^t$ the total green leaf area at growth cycle t and a constant Specifuc Blade Mass(e), and we get:

$$\mathcal{A}^t = \frac{1}{e} \sum_p \sum_{n=1}^{T_o} N^t_{b,p}(n) q^t_{b,p}(n)$$

The biomass production at growth cycle t is taken as a parameric function of $\mathcal{A}^t$ and of the environmental conditions. Inspired by Beer-Lambert's Law, a possible production function can be written:

$$u^t = E^t \beta \left(1 - \exp\left(-\gamma \mathcal{A}^t\right)\right) \tag{2.4}$$

where β and γ are empirical parameters

3 Propagation of uncertainity

It is usually not possible to determine explicitly the moments of the state vector(corresponding to biomass accumulation) for stochastic growth models. We face the problem of computing the statistics of a random variable which undergoes a non-linear transformation. It is useful to recall the following concepts.

Let x be a random vector in $\mathbb{R}^d$ with mean $\bar{x}$ and covariance matrix P. Let $g : \mathbb{R}^d \to \mathbb{R}$ a transformation of x, $y = g(x)$. Statistics on y can be obtained by Monte-Carlo simulations. However, the evaluation cost of g can be prohibitive for such simulations. We can obtain an approximation of the moments of $g(x)$ by considering the multivariate Taylor series expansion of $g(x)$ about$\bar{x}$.

$$g(x) = g(\bar{x}) + D_{\Delta x} g(\bar{x}) + \frac{1}{2!} D^2_{\Delta x} g(\bar{x}) + \cdots + \frac{1}{n!} D^n_{\Delta x} g(\bar{x}) + o(\|\Delta x\|^n) \tag{3.1}$$

where $D_{\Delta x}$ is the differential operator corresponding to the perturbation Δx

$$D^k_{\Delta x} = (\sum_{i=1}^{d} \Delta x_i \frac{\partial}{\partial x_i})^k$$

Since Δx is zero-mean variable, its first moment is zero and we have:

$$E[g(x)] = g(\bar{x}) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 g}{\partial x_i \partial x_j}(\bar{x}) E[\Delta x_i \Delta x_j] + \cdots \tag{3.2}$$

so we approximate:

$$E[g(x)] \approx g(\bar{x}) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 g}{\partial x_i \partial x_j}(\bar{x}) P_{ij} \tag{3.3}$$

For the variance we have:

$$Var[g(x)] = E[(g(x) - E[g(x)])^2]$$

with

$$g(x) - E[g(x)] = D_{\Delta x} g(\bar{x}) + \frac{1}{3!} D^3_{\Delta x} g(\bar{x}) + \cdots$$

Therefore, we have:

$$Var[g(x)] \approx \sum_{i,j} \frac{\partial g}{\partial x_i}(\bar{x}) \frac{\partial g}{\partial x_j}(\bar{x}) P_{ij} \tag{3.4}$$

If g is nota scalar and takes its value in $\mathbb{R}^m$, the above result generalizes to:

$$Cov(g_k(x), g_l(x)) \approx \sum_{i,j} \frac{\partial g_k}{\partial x_i}(\bar{x}) \frac{\partial g_l}{\partial x_j}(\bar{x}) P_{ij} \tag{3.5}$$

3.1 Description of the Discrete Dynamic System

The plant is modeled as a discrete dynamic system which grows at the start of discrete time steps(Growth Cycles). We consider an example of Sugar Beet, which has only 1 physiological age.

The state variable is chosen to be the vector of organ masses q^t, such that q^t defines the state of the system at growth cycle t.

$$q^t = \begin{pmatrix} q^t_{o_i,n_j} \\ \vdots \end{pmatrix}$$

where, $q^t_{o_i,n_j}$ denotes the mass of organ o_i of chronological age n_j at growth cycle t. The state of the system at each growth cycle can be expressed as a function of the state at the previous growth cycle, genetic parameters and some other exogenous variables.ie,

$$q^{t+1} = F(q^t, P, E)$$

In the GreenLab model, the following functional form is used:

$$q^{t+1}_{o,n+1} = F_{o,n}(q^t, P, E) = q^t_{o,n} + p^t_{o,n}\frac{u^t}{D^t} \tag{3.6}$$

where

$$u^t = E^t \mu S_p \left(1 - \exp\left(-\frac{k}{S_p}\sum_{i=0}^{t} N^t_{b,i} q^t_{b,i}\right)\right) \tag{3.7}$$

$$D^t = \sum_{o\in O, i=0}^{t} N^t_{o,i} p^t_{o,i} \tag{3.8}$$

We study the effect of variation in the following parameters:

- E^t(environment variable)
- μ
- S_p
- U_0(seed mass)

In order to obtain the mean and covariance of q^t, we use the Taylor series approximation as discussed above. The following recurrence relations are obtained:

$$\begin{aligned}
\overline{q^{t+1}_{o,n+1}} \quad = \quad & \overline{F_{o,n}} + \frac{1}{2}\sum_{i,j,k,l}\frac{\partial^2 F_{o,n}}{\partial q^t_{o_i,j}\partial q^t_{o_k,l}} Cov[q^t_{o_i,j}, q^t_{o_k,l}] + \sum_{i,j}\frac{\partial^2 F_{o,n}}{\partial q^t_{o_i,j}\partial E^t} Cov[q^t_{o_i,j}, E^t] \\
& + \sum_{i,j}\frac{\partial^2 F_{o,n}}{\partial q^t_{o_i,j}\partial \mu} Cov[q^t_{o_i,j}, \mu] + \sum_{i,j}\frac{\partial^2 F_{o,n}}{\partial q^t_{o_i,j}\partial S_p} Cov[q^t_{o_i,j}, S_p] + \sum_{i,j}\frac{\partial^2 F_{o,n}}{\partial q^t_{o_i,j}\partial U_0} Cov[q^t_{o_i,j}, U_0] \\
& + \frac{\partial^2 F_{o,n}}{\partial S_p \partial E^t}\sigma^2_{E^t} + \frac{\partial^2 F_{o,n}}{\partial S_p \partial S_p}\sigma^2_{S_p} + \frac{\partial^2 F_{o,n}}{\partial \mu \partial \mu}\sigma^2_{\mu} + \frac{\partial^2 F_{o,n}}{\partial U_0 \partial U_0}\sigma^2_{U_0}
\end{aligned}$$

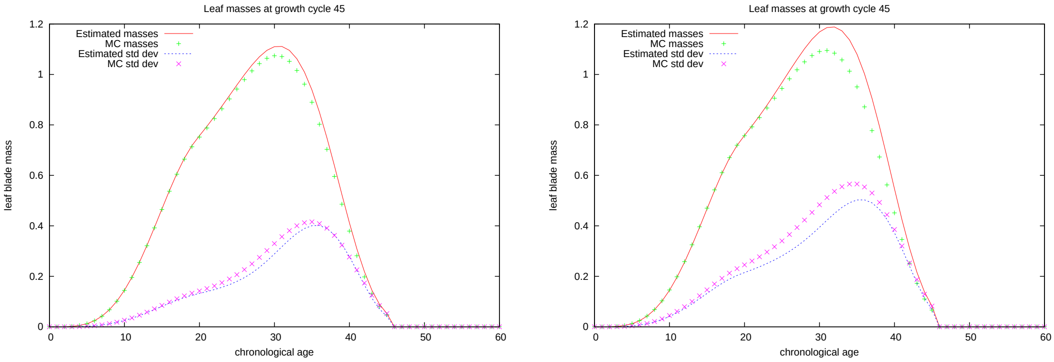
$$\begin{aligned}
Cov[q^{t+1}_{o_a,b+1}, q^{t+1}_{o_c,d+1}] \quad = \quad & \sum_{i,j,k,l}\frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_c,d}}{\partial q^t_{o_k,l}} Cov[q^t_{o_i,j}, q^t_{o_k,l}] \\
& + \sum_{i,j}\frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_c,d}}{\partial E^t} Cov[q^t_{o_i,j}, E^t] + \sum_{i,j}\frac{\partial F_{o_c,d}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_a,b}}{\partial E^t} Cov[q^t_{o_i,j}, E^t] \\
& + \sum_{i,j}\frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_c,d}}{\partial \mu} Cov[q^t_{o_i,j}, \mu] + \sum_{i,j}\frac{\partial F_{o_c,d}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_a,b}}{\partial \mu} Cov[q^t_{o_i,j}, \mu] \\
& + \sum_{i,j}\frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_c,d}}{\partial S_p} Cov[q^t_{o_i,j}, S_p] + \sum_{i,j}\frac{\partial F_{o_c,d}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_a,b}}{\partial S_p} Cov[q^t_{o_i,j}, S_p] \\
& + \sum_{i,j}\frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_c,d}}{\partial U_0} Cov[q^t_{o_i,j}, U_0] + \sum_{i,j}\frac{\partial F_{o_c,d}}{\partial q^t_{o_i,j}}\frac{\partial F_{o_a,b}}{\partial U_0} Cov[q^t_{o_i,j}, U_0] \\
& + \frac{\partial F_{o_a,b}}{\partial E^t}\frac{\partial F_{o_c,d}}{\partial E^t}\sigma^2_{E^t} + \frac{\partial F_{o_a,b}}{\partial \mu}\frac{\partial F_{o_c,d}}{\partial \mu}\sigma^2_{\mu} + \frac{\partial F_{o_a,b}}{\partial S_p}\frac{\partial F_{o_c,d}}{\partial S_p}\sigma^2_{S_p} + \frac{\partial F_{o_a,b}}{\partial U_0}\frac{\partial F_{o_c,d}}{\partial U_0}\sigma^2_{U_0}
\end{aligned}$$

$$Cov[q^{t+1}_{o_a,b+1}, E^{t+1}] = \sum_{i,j} \frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}} Cov[q^t_{o_i,j}, E^t] + \frac{\partial F_{o_a,b}}{\partial E^t} \sigma^2_{E^t}$$

$$Cov[q^{t+1}_{o_a,b+1}, \mu] = \sum_{i,j} \frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}} Cov[q^t_{o_i,j}, \mu] + \frac{\partial F_{o_a,b}}{\partial \mu} \sigma^2_{\mu}$$

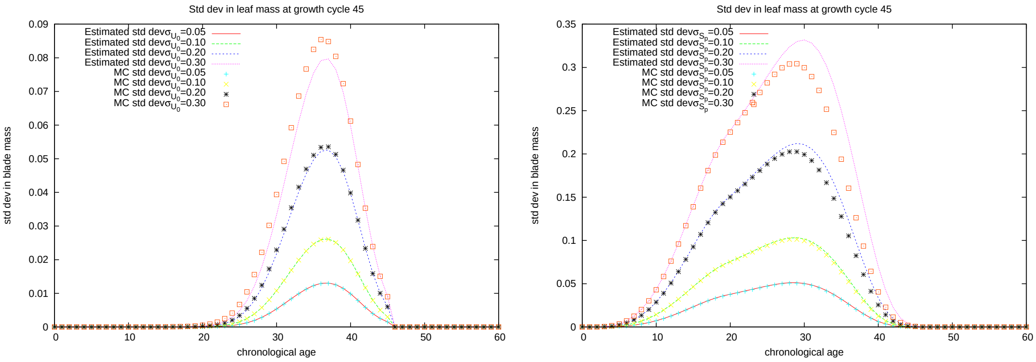
$$Cov[q^{t+1}_{o_a,b+1}, S_p] = \sum_{i,j} \frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}} Cov[q^t_{o_i,j}, S_p] + \frac{\partial F_{o_a,b}}{\partial S_p} \sigma^2_{S_p}$$

$$Cov[q^{t+1}_{o_a,b+1}, U_0] = \sum_{i,j} \frac{\partial F_{o_a,b}}{\partial q^t_{o_i,j}} Cov[q^t_{o_i,j}, U_0] + \frac{\partial F_{o_a,b}}{\partial U_0} \sigma^2_{U_0}$$



(a) $\sigma_E = 0.1$,$\sigma_\mu = 0.1$, $\sigma_{S_p} = 0.1$, $\sigma_{U_0} = 0.1$ (b) $\sigma_E = 0.2$, $\sigma_{S_p} = 0.2$, $\sigma_{U_0} = 0.2$

Figure 1: Mean and standard deviation in leaf blade mass



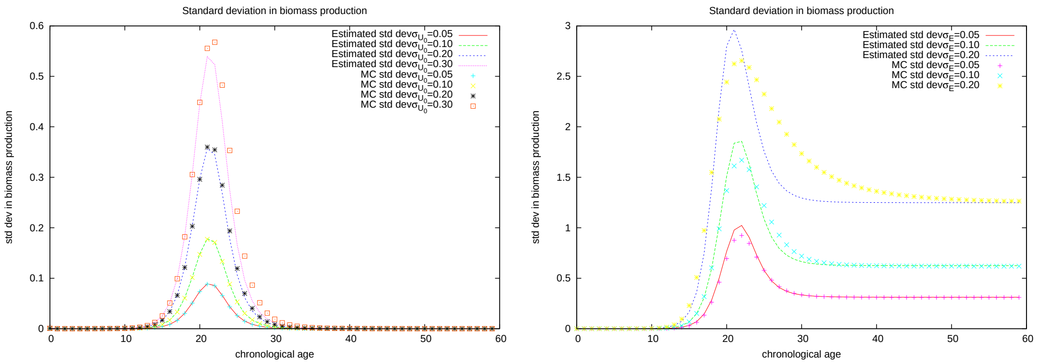
(a) Variance in U_0 (b) Variance in S_p

Figure 2: Mean and standard deviation in leaf blade mass

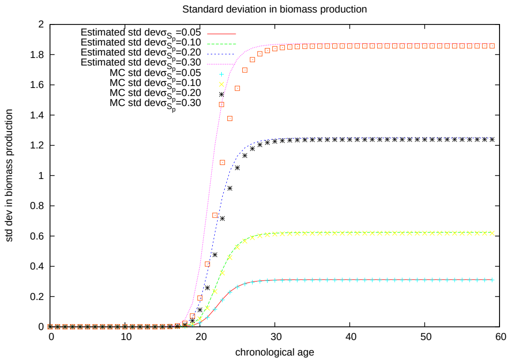
Monte-Carlo simulations were done and compared with the approximation results obtained using the above equations. Figure 1 and 2 show the comparisons of mean and variance for different values of parameter variance.(All values of the standard deviation in parameters are expressed as fractions of their means. The mass of dead leaves has also been included).

3.2 Effect on biomass production

Biomass produced at growth cycle t, u_t is a direct function of the state variable q_t and other parameters. The effect of variance in parameters was observed on the mean and standard deviation of u_t. The results are shown in Figure 3



(a) variance in U_0 (b) variance in E



(c) variance in S_p

Figure 3: Standard deviation of biomass production(u_t) with variance in different parameters

3.3 Variance in phyllochron

Phyllochron is a measure of the thermal time received by the plant between successive emergence of new organs. Due to variation in the rhythm of leaf appearance the phyllochron for each plant in a population varies. However, the sink functions vary with chronological age(and not with thermal time). In order to model the effects of variation in phyllochron, we changed the above model as follows(r is the length of the phyllochron measured in timecycles, the length of the timecycle corresponds to the rate of leaf appearace for the fastest growing plant):

- n, which earlier represented chronological age, now represents the rank of the organ in order of appearance.
- $E[t]$ is replaced by $rE[t]$ since the environmental input increases by a factor of r
- The sink functions

  $$p^t_{o,p}(n) = P_{o,p}\left(\frac{n+0.5}{T^i_{o,p}-T^f_{o,p}}\right)^{\alpha_o}\left(1-\frac{n+0.5}{T^i_{o,p}-T^f_{o,p}}\right)^{\beta_o}\frac{1}{max_{x\in[0,1]}x^{\alpha_o}(1-x)^{\beta_o}}$$

  are replaced by

  $$p^t_{o,p}(n) = P_{o,p}\left(\frac{r(n+0.5)}{T^i_{o,p}-T^f_{o,p}}\right)^{\alpha_o}\left(1-\frac{r(n+0.5)}{T^i_{o,p}-T^f_{o,p}}\right)^{\beta_o}\frac{1}{max_{x\in[0,1]}x^{\alpha_o}(1-x)^{\beta_o}}$$

The earlier version hence corresponds to r=1.

4 Parameter Estimation

We consider a dynamic system described by a discrete time model. Let $(t_n)_{n\in[0,N]}$ be a finite sequence of successive times corresponding to the evolution steps. We denote by $X_n \in \mathbb{R}^x$ the vector of state variables at t_n, $U_n \in \mathbb{R}^u$, the set of exogenous variables(entries, controls,...) at t_n, and $P \in \mathbb{R}^p$, the vector of model parameters. The growth model can thus be expressed as

$$X_{n+1} = F_n(X_n, U_n, P) \tag{4.1}$$

with X_0 given. Let $Y \in \mathbb{R}^y$ be a vector of experimental observations made on the system at a given time t_N. These observations correspond to model outputs $tildeY$ deduced from X_N.

$$\tilde{Y} = G(X_N, P) \tag{4.2}$$

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

$$\epsilon(P) = Y - \tilde{Y}(P) \tag{4.3}$$

4.1 Maximum likelihood estimator

The likelihood of Y as a function of P,$\prod(Y|P)$, is the probablity density of Y when P is the vector of model parameters. Identification of the system parameters can be made by finding $\hat{P}$ that maximizes $\prod(Y|P)$. We suppose that ϵ is a gaussian vetor of covariance matrix Σ. We have:

$$\prod(Y|P) = [(2\pi)^y \det(\Sigma)]^{-1/2} \exp[-\frac{1}{2}(Y - \tilde{(Y)}(P))^T \Sigma^{-1} (Y - \tilde{(Y)}(P))] \tag{4.4}$$

If Σ is known, $\hat{P}$ is given by the Gauss-Markov estimator, that is to say by minimizing the quadraic criteria:

$$\hat{P} = Argmin_P((Y - \tilde{(Y)}(P))^T \Sigma^{-1} (Y - \tilde{(Y)}(P))). \tag{4.5}$$

However, the covariance matrix Σ is generally unknown. In such cases, estimators of Σ and $\hat{P}$ can be derived if we have a large number of repetitions of the experimental observations. It is not possible to have that since for plants, repetitions cause strong variability between individuals.

The classical way to overcome this problem is to model the error covariance. The simplest possibility o sto suppose that output errors are independent and homoscedastic, that is to say $\Sigma = \sigma^2 I$, wher I is the identity matrix. The maximum likelihood estimator is thus equivalent to the ordinary least-square estimator.

4.2 Heteroscedasticity

The homoscedastic hypothesis is not a plausible assumption for the GreenLab model since the masses in Y correspond to different types of organs with very different size orders. We thus consider a more general hypothesis, corresponding to an error model studies in detail by Taylor(1977) [8]: the y observations are classified into q groups, with each of the y_i error terms in group i having common unknown variance θ_i $(1 \le i \le q)$ and the errors supposedly mutually independent. For the observation vector on plants, each group corresponds toa type of organ. We suppose:

$$\epsilon(P) = Y - \tilde{Y}(P) = \mathcal{N}(0, \Sigma). \tag{4.6}$$

with Σ a diagonal matrix of rank y :

$$\Sigma = \begin{pmatrix} \theta_1 I_{y_1} & 0 & 0 & \cdots & 0 \\ 0 & \theta_2 I_{y_2} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \theta_{q-1} I_{y_{q-1}} & 0 \\ 0 & \cdots & 0 & 0 & \theta_q I_{y_q} \end{pmatrix} \tag{4.7}$$

with I_k, the identity matrix of order k.

4.3 Two stage Aitken estimator

For such type of systems, a two stage Aitken estimator can be used(Taylor(1977) [8]). 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 with an estimated covariance matrix $\hat{\Sigma}$:

$$\hat{P}_{2SA} = Argmin_P((Y - \tilde{Y}(P))^T \tilde{\Sigma}^{-1}(Y - \tilde{Y}(P))). \tag{4.8}$$

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

$$\hat{\theta}_i = \frac{1}{y_i - p}(Y_i - \tilde{Y}_i(\hat{P}_{LS,i}))^T (Y_i - \tilde{Y}_i(\hat{P}_{LS,i})) \tag{4.9}$$

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

$$\hat{P}_{LS,i} = Argmin_P((Y_i - \tilde{Y}_i(P))^T (Y_i - \tilde{Y}_i(P))) \tag{4.10}$$

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

$$L(P, \theta) = \frac{y}{2}2\pi + \sum_{i=1}^{q} \frac{y_i}{2} \ln(\theta_i) + \frac{1}{2} \sum_{i=1}^{q} \frac{1}{\theta_i} \left(\left(Y_i - \tilde{Y}_i(P) \right)^T \Sigma^{-1} \left(Y_i - \tilde{Y}_i(P) \right) \right) \tag{4.11}$$

so that the maximum likelihood estimators for P and θ are:

$$\hat{\theta}_i = \frac{1}{y_i} \left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right); 1 \leq i \leq q, \tag{4.12}$$

$$\hat{P} = Argmin_P \left(\sum_{i=1}^{q} \frac{1}{\theta_i} \left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right) \right) \tag{4.13}$$

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

$$\sum_{i=1}^{q} \frac{y_i}{\left(Y_i - \tilde{Y}_i(P) \right)^T \left(Y_i - \tilde{Y}_i(P) \right)} \left(\frac{\partial \tilde{Y}}{\partial P} \left(\hat{P} \right) \right)^T \left(Y_i - \tilde{Y}_i(\hat{P}) \right) = 0. \tag{4.14}$$

We attempt to solve this iteratively starting with

$$\hat{\theta}_i = \frac{1}{y_i} \left(Y_i - \tilde{Y}_i(\hat{P}_{LS,i}) \right)^T \left(Y_i - \tilde{Y}_i(\hat{P}_{LS,i}) \right) \tag{4.15}$$

then deducing $\hat{P}$ and again correcting $\hat{\theta}$ with this new version of $\hat{P}$ till a convergence criterion is reached.

4.4 Multi-Fitting

In order to take into account the dynamics of plant growth, it is interesting to consider an observation vector composed of several intermediate observations at different growth stages $(\tau_i)_{1 \leq i \leq k}$,

$$Y = (Y^{\tau_1}, \cdots, Y^{\tau_k})^T$$

and the corresponding model outputs

$$\tilde{Y} = (G(X^{\tau_1}, P), \cdots, G(X^{\tau_k}, P))^T$$

The same heteroscedasticity hypothesis can be done, considering a constant error variance for each group of organs at each observation time.

5 Conclusions and Results

The estimated organ masses and their variances in the plant population agree with the Monte-Carlo simulation results. The deviation between the two increases with the value of the variances, which is to be expected. The work on parameter estimation is not yet completed and is being worked on. A basic understanding of the problem has been done as described above. The work done on error propagation and modeling will be presented at PMA'09, the Third International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications, Beijing, China.

References

[1] Aristid Lindenmayer. Mathematical models for cellular interactions in development i. filaments with one-sided inputs. *Journal of Theoretical Biology*, 18(3):280–299, March 1968.

[2] L. Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *Automatic Control, IEEE Transactions on*, 24(1):36–50, Feb 1979.

[3] Lennart Ljung and Torsten Soderstrom. *Theory and Practice of Recursive Identification (Signal Processing, Optimization, and Control).* The MIT Press, October 1983.

[4] P. Prusinkiewicz and Aristid Lindenmayer. *The algorithmic beauty of plants.* Springer-Verlag New York, Inc., New York, NY, USA, 1990.

[5] Przemyslaw Prusinkiewicz, Aristid Lindenmayer, and James Hanan. Development models of herbaceous plants for computer imagery purposes. In *SIGGRAPH '88: Proceedings of the 15th annual conference on Computer graphics and interactive techniques*, pages 141–150, New York, NY, USA, 1988. ACM.

[6] Grzegorz Rozenberg and Arto. Salomaa. The mathematical theory of l systems / grzegorz rozenberg, arto salomaa, 1980.

[7] Alvy Ray Smith. Plants, fractals, and formal languages. *SIGGRAPH Comput. Graph.*, 18(3):1–10, 1984.

[8] William E Taylor. Small sample properties of a class of two stage aitken estimators. *Econometrica*, 45(2):497–508, March 1977.

[9] Hongping Yan, Philippe de Reffye, Chunhong Pan, and Bao-Gang Hu. Fast construction of plant architectural models based on substructure decomposition. *J. Comput. Sci. Technol.*, 18(6):780–787, 2003.