HAL
open science

Plants as Combinatorial Structures and Applications

Cedric Loi, Paul-Henry Cournède, Jean Françon

To cite this version:

Cedric Loi, Paul-Henry Cournède, Jean Françon. Plants as Combinatorial Structures and Applications. Third International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA09), Nov 2009, Pékin, China. pp.319-326, 10.1109/pma.2009.27 . inria-00546283

HAL Id: inria-00546283

https://inria.hal.science/inria-00546283v1

Submitted on 14 Dec 2010

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.

Plants as Combinatorial Structures and Applications

Cédric Loi
Ecole Centrale de Paris, France
cedric.loi@ecp.fr

Paul-Henry Cournède
Ecole Centrale de Paris, France
paul-henry.cournede@ecp.fr

Jean Françon
Université de Strasbourg, France
jean_francon@hotmail.fr

Abstract

In this article, we introduce a new method allowing the computation of the distribution associated to the number of complex structures in plants deriving from GreenLab type growth model. In order to use this method, we set a new mathematical framework based on combinatorics. We show how plants can be seen as plane rooted trees and how their topology can be described by Dyck words. Moreover, we integrate plant growth in the formalism by adapting stochastic F0L-systems to the framework. This new representation enables not only the computation of the distribution associated to all types of organ but also the analysis of particular patterns thanks to a symbolic method. This approach gives new kinds of applications such as estimating the age of a plant from a set of botanical data.

1. Introduction

Models of plant development describe the dynamic creation of organs (leaves, internodes, flowers/fruits) and how they organize themselves to form plant structures. When the smallest scale of interest is that of organs (and not cells), discrete models such as *GreenLab* (see [1] and [2]) are generally used to simulate plant structural development. Several formalisms relying on language theory (see [3] and [4]) have been proposed to describe the plant topology deriving from these models. Such formalisms integrate easily growth dynamics by using L-systems. Methods based on generating functions enable us to compute the moments and the distribution of the number of organs in a plant (see [4]). However, these methods work only with certain types of organ and fail when regarding complex structures (apexes, a particular sequence of metamers,. . .). Other formalisms relying on combinatorial concepts have been set for computer graphics needs (see [5]). Such formalisms enable a wider description of the topology and the development. However, none of them have completely taken advantage of all the benefits of combinatorics.

The aim of this article is to build a new method allowing the computation of the distribution associated to the number of complex structures for GreenLab type plant growth models. To that purpose, we set a combinatorial framework integrating the previous formalisms in a more general sense and we use it through fundamental tools of combinatorics such as the symbolic method. This method enables us to count the number of plant having a specific topology and to determine not only the distributions of all kinds of organs (metamers, fruit, . . .) but also the distributions associated to particular structures and patterns in a plant.

First, basic botanical concepts of discrete plant growth model like Greenlab are recalled. Then, a new combinatorial framework is set and describes plants as plane rooted trees. The topology is represented by Dyck words and the development by stochastic F0L-systems. Such formalism enables us to use classical tools of combinatorics and a new symbolic method is introduced. In the third section, we show how to adapt this method to study plants. Finally, the last section shows the benefits of combinatorics regarding plants through two examples of applications.

2. Botanical concepts to model plant development

We recall some basic concepts concerning models of plant development like GreenLab (see [4] for more details on the botanical modelling). GreenLab is a discrete time model whose time step is called *Growth Cycle*. Each plant is composed of a succession of elementary units called *metamers*. A *Growth Unit* is the set of metamers built by a bud during a growth cycle. The *Chronological Age* (= CA) of a plant (or of an organ) is defined as the number of growth cycles it has existed for. Concerning the architecture of the plant, the axis can be listed into different categories depending on their morphological parameters. Thus, as explained in [6], the concept of *Physiological Age* (= PA) was introduced to represent the different types of axes. Let P be the maximal PA (in general, $P \leq 5$). It corresponds to the ultimate state of differentiation for an axis, it is usually short, without branches. The apical meristem or bud of an axis is thus characterized by the PA of the growth unit that it may produce and a metamer is characterized by its PA i (which is the PA of the growth unit that it belongs to).

3. Plants as combinatorial structures

The aim of this section is to set a combinatorial framework to describe plant topology.

3.1. Describing a plant with a plane rooted tree

Let us first define some basic notions of combinatorics (see [7], [8]). A *linear graph* is a collection of points (called *nodes*) and a collection of lines (called *edges*) which describes the connections of points. A graph is *connected* if every pair of nodes is joined by a path, that is, a collection of lines of the form p_1p_2,p_2p_3,. . .,$p_{k-1}p_k$ with all points p_1 to p_k distinct. A *cycle* is a path starting and ending by the same point. Then, a tree is a connected linear graph without cycles.

If there exists one node that can be distinguished from the others, then the tree is *rooted* and the node is called the root of the tree. Figure 1 is a collection of trees drawn with the root below. A *plane tree* is defined as a tree in which subtrees

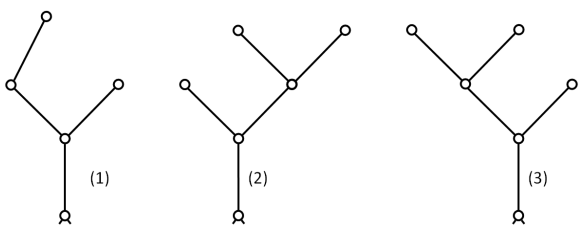

Figure 1. Examples of rooted trees

coming from a common node are ordered and represented from left to right. Thus, the representations (2) and (3) of Figure 1 are equivalent as rooted trees, but they become distinct objects when regarded as plane trees.

To make the connections between plants and plane rooted trees (= PRT), we have to make some hypotheses:

- We are only interested in the topology of the plant.
- Metamers are reduced to their internodes.
- A metamer is represented by an edge and ends with a node at each extremity.
- Two connected metamers are represented by two edges that have a common node.

Then, a plant is a PRT (see Figure 2). A seed is a PRT with only one node. Characteristics can be added to nodes

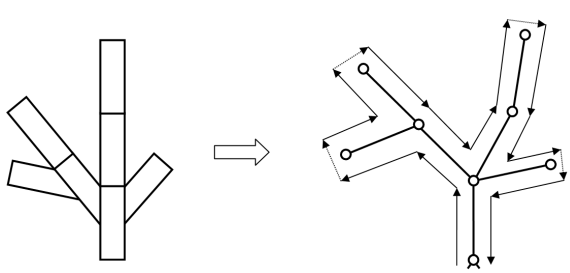

Figure 2. Representation of a plant by a plane rooted tree

and edges. Then, the tree is said to be *labelled*. Important botanical data can be taken into account such as Physiological Age (= PA). Every edge is labelled by the PA of the metamer it represents. A node is labelled by the PA of the metamer above it. The PA of the root is 1. Figure 3 gives an example.

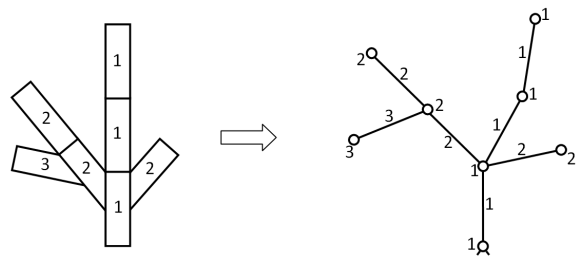

Figure 3. Physiological age and labelled trees

N.B.: in this article, we consider only 2D plants. A 3D plant can be represented by several 2D plants. However nothing changes from a probability point of view if we are careful regarding parameters identification.

3.2. Coding a plant by a word

The topology of a plant can be represented by a word. Several ways of description may be used. The most basic one is that of Dyck words (see [9]). Every PRT is described by a Dyck word on the alphabet $V_1 = \{z, z'\}$ as follows:

- each element of the tree is visited according to the prefix order.
- an edge visited from bottom to top is represented by the letter z.
- an edge visited from top to bottom is represented by the letter z'.

A PRT is visited according to the *prefix order* (see [7]) if it is traversed starting from the root, proceeding left-to-right to visit each subtree, and backtracking upwards once a subtree has been completely traversed. Figure 2 gives an example of prefix order. The Dyck word associated to the right PRT of Figure 2 is thus $w_1 = zzzz'zz'z'zzz'z'zz'z'$. This representation is very simple and more complicated codes can be introduced to describe the structure of a tree more precisely. In the case of labelled trees, the letters can be indexed by the value of one label. For instance, the alphabet $V_2 = \{z_1, z_1', z_2, z_2', \ldots, z_P, z_P'\}$ may be used to describe the plant of Figure 3. The associated Dyck word (in its extended sense) is then $w_2 = z_1z_2z_3z_3'z_2z_2'z_2'z_1z_1z_1'z_1'z_2z_2'z_1'$. Finally, you can add special symbols to the alphabet to describe particular elements of the plant. For example, the letter l can represent a leaf and the letter f a fruit. If you are interested in the number of apexes (= terminal nodes) of the plant, you can add the letter a after every sequence of type zz'. The associated word for the PRT

of Figure 3 taking into account apexes is thus $w_3 = z_1z_2z_3z_3'az_2z_2'az_2'z_1z_1z_1'az_1'z_2z_2'az_1'$ and the associated alphabet is $V_3 = V_2 \cup \{a\}$.

N.B.:

1) w_1, w_2 and w_3 describe the same PRT (see Figure 3) with different levels of details (they are not built on the same alphabet). w_3 is more precise than w_2 and w_2 is more precise than w_1. Notice that we have $V_1 \subset V_2 \subset V_3$. The level of details varies with the size of the alphabet. Depending of the needs, we will choose one of them.
2) For a given alphabet, a unique word is associated to a PRT. The converse is true.

In the sequel, the set of all PRTs is denoted by $\mathcal{T}$. Let V be an alphabet and v its cardinal. Then, for all $t \in \mathcal{T}$, w_t is the word associated to t built on V. Finally, let $W = \{w_t, t \in \mathcal{T}\}$ be the set of all words built on V.

3.3. Growth dynamics and stochastic F0L-systems

Up to now, we have described plant topology but we have not taken into account the dynamics of plant development. We are interested in the evolution of the structure of a plant with its CA starting with a seed (or a bud). This evolution can be described by a sequence $(t^n)_{n\in\mathbb{N}}$ of $\mathcal{T}$ (or the corresponding sequence $(w^n)_{n\in\mathbb{N}}$ of W). The evolution from t^n to t^{n+1} (or from w^n to w^{n+1}) is given by a set of rules that can be deterministic or stochastic.

Stochastic F0L-systems are well suited to describe these rules (see [10], [11] and [12]). They are generative parallel rewriting grammars whose production rules are associated to a set of distribution probabilities. The alphabet V_L associated to a stochastic F0L-system L contain terminal elements and nonterminal elements. A stochastic F0L-system generates a sequence of words $(w^n)_{n\in\mathbb{N}}$ built on this alphabet. We get w^{n+1} by replacing all the nonterminal elements x of w^n by a sequence of letters y according to the production rules of L. There may be several possible evolutions y for a letter x. In that case, a probability $\pi_{x,y}$ is associated to each evolution. Therefore a stochastic F0L-system (see [4]) can be simply defined a construct $L = \langle A, \pi\rangle$ where:

- $A \subset W_L$ is called the set of axioms of L. The axiom is the word initiating the sequence generated by L (= w^0). It can be any element of A.
- π is a transition matrix from V_L to W_L. For every $(x, y) \in V_L \times W_L$, the component $\pi_{x,y}$ is the probability for the letter x to evolve into the word y.

N.B.: Most of the time, the production rules of a stochastic F0L-system are represented by a set of stochastic automata (see Figure 4 and 5).

Bud activity can be represented by a set of stochastic automata (see [13]). Therefore, the growth dynamics of plants can be represented by a stochastic F0L-system L. In Section 3.2 plants were only described by static elements (metamers, fruits, ...). Thus, if we want to take into account the growth dynamics due to bud activity, the alphabet V needs to be completed with a set of nonterminal elements S which represents buds. Depending of the level of details we want, we can have $S = \{s\}$ (for simple code), $S = \{s_1, \ldots, s_P\}$ (for labelled trees with PAs) or something more elaborate. Let us set $V_L = V \cup S$ and let $W_L = \{w_t, t \in \mathcal{T}\}$ be the set of all words built on V_L. Thus, the evolution of the structure of a plant is represented by a sequence $(w_L^n)_{n\in\mathbb{N}}$ of W_L. w_L^0 is generally a bud of PA 1. The set of axioms A of L is S. The components of the transition matrix π give the probabilities of all possible evolutions of buds according to the stochastic automata. Each term w_L^n contains at least symbols for the structure (*i.e.* metamers) and symbols for the growth dynamics (*i.e.* buds). We get w_L^{n+1} by replacing every letter s ($\in S$) of w_L^n by a word y ($\in W_L$) with a probability $\pi_{s,y}$. By doing so, we are able to build cycle after cycle all the terms of the sequence $(w_L^n)_{n\in\mathbb{N}}$ (which is equivalent to build the associated sequence of PRTs $(t^n)_{n\in\mathbb{N}}$ of $\mathcal{T}$).

In case of deterministic growth, all the components of π are equal to either 0 or 1. Figure 4 gives an example of deterministic growth. After 2 cycles, the plant of Figure 4 is

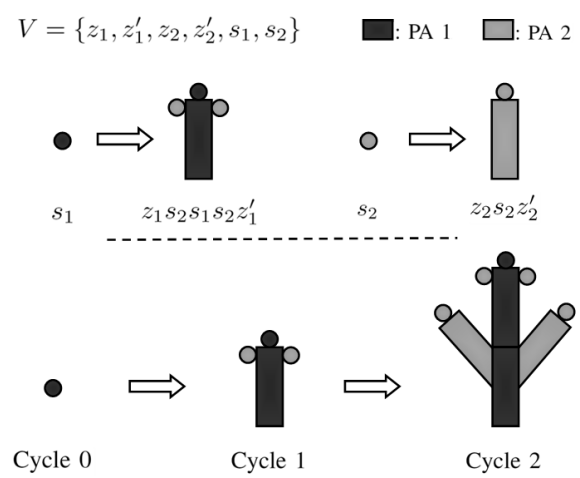

Figure 4. Deterministic growth and stochastic F0L-system

described by the word $w_L^2 = z_1z_2s_2z_2'z_1s_2s_1s_2z_1'z_2s_2z_2'z_1'$. We have $w_L^0 = s_1$ and $w_L^1 = z_1s_2s_1s_2z_1'$.

As for stochastic growth, buds can evolve into different states according to a probability distribution (see Figure 5). The plant of Figure 3 is one possible occurence of the stochastic F0L-system of Figure 5. In the sequel, $L = \langle S, \pi\rangle$ will denote a stochastic F0L-system and $\mathcal{T}_L^n[s]$ ($\subset \mathcal{T}$) the set of all PRTs generated by L with $s \in S$ as axiom after n growth cycles ($n \in \mathbb{N}$). For all $t \in \mathcal{T}_L^n[s]$, P_t denotes the

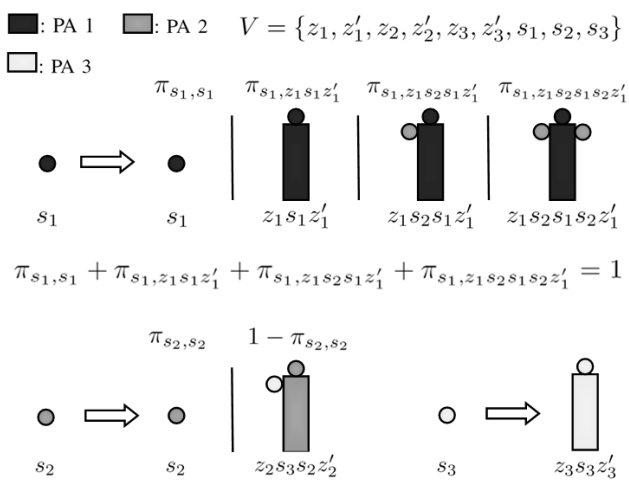

Figure 5. Stochastic growth and stochastic F0L-system

probability to get the PRT t by using the production rules of L.

4. The symbolic method

The symbolic method is a very effective method to analyse combinatorial structures and, as a consequence, plays an important role in analytic combinatorics (see [7] for more details). Concerning plants, this method enables us to count the number of structures with a given characteristic (see Section 4.1) or to find the distribution of the number of organs for dynamic plant growth models such as GreenLab (see Section 4.2).

4.1. Counting plane rooted trees

In this section, we are only interested in plant topology without growth dynamics. The aim is to find the number of PRTs with particular characteristics (for example, the number of plants with 7 metamers). To do so, we have to introduce new concepts: the *counting function* and the *ordinary generating function* (= OGF).

Let $V = \{u_1, \ldots, u_v\}$ be an alphabet and $v = Card(V)$. The counting function c is a map from $\mathcal{T} \times V$ to $\mathbb{N}$ such that, for all $(t, u) \in \mathcal{T} \times V$, $c(t, u)$ gives the number of letters u in the word w_t. Let T be a subset of $\mathcal{T}$. The ordinary generating function associated to T on V is a map from $[0, 1]^v$ to $\mathbb{N}$ defined as:

$$\Phi(u_1, \ldots, u_v) = \sum_{t \in T} |w_t| \quad \text{with} \quad |w_t| = \prod_{i=1}^{v} u_i^{c(t,u_i)}$$

By reordering the terms of the sum, we get:

$$\Phi(u_1, \ldots, u_v) = \sum_{t \in T} |w_t| = \sum_{(n_1,\ldots,n_v) \in \mathbb{N}^v} C_{n_1,\ldots,n_v} \prod_{i=1}^{v} u_i^{n_i} \tag{1}$$

$C_{n_1,\ldots,n_v}$ is the number of PRTs of T having n_i letters u_i in their word. As a consequence, to determine the OGF of T on V is equivalent to find the number of trees with particular characteristics. However, most of the time, the computation of the OGF Φ is difficult. Thus, the idea is to use the *structural property* of PRTs to get a set of functional equations with Φ as variable: a PRT is either a node or a node connected to one or several PRTs. The symbolic method for OGFs is thus the following:

1) Identify the structure of the plant.
2) Find the appropriate alphabet V and the subset T.
3) Write the associated OGF Φ.
4) Use the structural property of PRTs restricted to T to get a set of functional equations with Φ as variable.
5) Use these equations to compute the coefficients of the OGF.

N.B.: the number of variables in a OGF is equal to the cardinal of V. However, most of the time, we are not interested in a complete description of a plant but only in particular characteristics. In that case, OGFs can be simplified by wiping out variables we are not interested in. For instance, for all $t \in \mathcal{T}$, we have $c(t, z) = c(t, z')$. Therefore, the symbols z and z' contain the same information regarding counting. As a consequence, we do not need to insert the symbol z' in the OGF Φ of T on V. Thus, in the sequel, z' will not be a variable of OGFs. OGFs with only elements of interest as variables are called simplified OGFs. In the third step of the symbolic method, OGFs can be replaced by simplified OGFs.

At step 4, you need at least as many equations as the number of possible labels for the root to describe completely the topology of the plant. Let us process this method throughout an example. The aim is to find the number of full binary trees having n edges with $n \in \mathbb{N}$. Full binary trees (= FBT) are PRTs such that each node has either 0 or 2 downstream nodes. T is thus the set of all FBTs. The alphabet is only $V = \{z, z'\}$. Then, for all $z \in [0, 1]$, the simplified OGF is:

$$\Phi(z) = \sum_{t \in T} z^{c(t,z)} = \sum_{n \in \mathbb{N}} C_n z^n$$

with C_n the number of FBTs with n edges. The structural property restricted to T is: a FBT t is either a node or a node connected to two FBTs t_1 and t_2. In the last case, we have $c(t, z) = 2 + c(t_1, z) + c(t_2, z)$. Then:

$$\Phi(z) = \sum_{t \in T} z^{c(t,z)} = 1 * z^0 + \sum_{(t_1,t_2) \in T^2} z^{2+c(t_1,z)+c(t_2,z)}$$

$$= 1 + z^2 * \left(\sum_{t \in T} z^{c(t,z)} \right)^2 = 1 + z^2 * (\Phi(z))^2$$

The only possible solution of this equation of degree 2 is:

$$\Phi(z) = \tfrac{1-\sqrt{1-4z^2}}{2z^2} = \sum_{n \in \mathbb{N}} \frac{1}{n+1} \binom{2n}{n} z^{2n}$$

We deduce $C_{2n} = \frac{1}{n+1}\binom{2n}{n}$ and $C_{2n+1} = 0$ for $n \in \mathbb{N}$.

4.2. Finding distributions in dynamical plant growth models

In this section, we use the symbolic method on plant models taking into account growth dynamics such as GreenLab. The structure of PRTs is led to evolve according to a set of stochastic rules (see Section 3.3). The aim is to determine the distribution of elements of interest in particular models of plant growth. The counting function introduced in Section 4.1 plays an important role here too. OGFs are replaced by *probability generating functions* (= PGFs).

Let $V = \{u_1, \ldots, u_v\}$ be an alphabet and $v = Card(V)$. Let $L = \langle S, \pi \rangle$ be a stochastic F0L-system. For $n \in \mathbb{N}$ and $s \in S$, the probability generating function associated to $\mathcal{T}_L^n[s]$ on V is a map from $[0,1]^v$ to $[0,1]$ defined as:

$$\Psi^n[s](u_1, \ldots, u_v) = \sum_{t \in \mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s}]} P_t \, |w_t| \quad \text{with} \quad |w_t| = \prod_{i=1}^{v} u_i^{c(t,u_i)}$$

where P_t is the probability of occurence of w_t (see Section 3.3). By reordering the terms of the sum, we get:

$$\Psi^n[s](u_1, \ldots, u_v) = \sum_{t \in \mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s}]} P_t \, |w_t| = \sum_{(k_1,\ldots,k_v) \in \mathbb{N}^v} P_{k_1,\ldots,k_v}^{n,s} \prod_{i=1}^{v} u_i^{k_i} \tag{2}$$

N.B.: in the same way as in Section 4.1, elements of PGFs which do not interest us can be wiped out. For instance, most of the time, we are not interested in the distribution of the number of buds. Therefore, in that case, PGFs will not have any element of S as variables. Thus, PGFs are said simplified.

$P_{k_1,\ldots,k_v}^{n,s}$ is the probability for a plant of CA n with s as root to have k_i elements u_i for $i \in \{1, \ldots, v\}$. As in Section 4.1, to determine the PGF of $\mathcal{T}_L^n[s]$ on V with $s \in S$ is equivalent to find the distribution of the number of organs of interest in particular plant structures. In the same way as in Section 4.1, we can write a symbolic method for dynamic growth models:

1) Identify the structure of the plant and all stochastic behaviours.
2) Deduce the associated stochastic automata.
3) Identify all the elements you need to describe the stochastic growth and find the appropriate alphabet V.
4) Write the stochastic F0L-system $L = \langle S, \pi \rangle$.
5) Identify the elements or structures whose distributions you want to compute.
6) Write the associated simplified PGFs $\Psi^n[s]$ for $n \in \mathbb{N}$ and $s \in S$.
7) Use the structural property to get a set of recursive equations between Ψ^n and Ψ^{n+1}.
8) Use these equations to compute the coefficients of the PGFs.

At step 4, the transition matrix π is determined by using the stochastic automata of step 2. At step 5, elements can be any organs and structures refer to particular patterns (*i.e.* a set of consecutive letters) in Dyck words. At step 6, the structural property is generally more complex than that of Section 4.1. Let us consider a plant of CA n. To describe completely the topology of the plant, you need at least as many equations as the number of possible labels for the root. For each of these equations, you have to break down the structure of the PRT in a way which highlights how the elements of interest are laid out in this PRT and how they are connected to subPRTs of CA $n-1$. Section 5.1 gives several examples of how to proceed.

N.B.: when we are interested in permanent elements or structures, this method is completely equivalent to the method relying on composition of generating functions (see [4]). It will appear clearly in Section 5.1 where we show how to connect the two methods. However, the method of composition of generating functions may not work when the elements of interest depend on a structure that may evolve (for example transitory patterns in words). For instance, in Section 5.2, we are interested in the distribution of the number of apexes in the plant. Apexes can not be described by stochastic F0L-systems because their evolution is coupled to that of other elements in the plant. For that example, only the symbolic method can work.

5. Applications

This section illustrates the use of the symbolic method throughout two examples.

5.1. Finding distributions and moments in stochastic plant growth model

An immediate application of the symbolic method is to determine the distribution and the moments of the number of organs in a particular plant structure. In this section, we are interested in the distribution and the moments of the number of fruits. The stochastic automata are given by Figure 6. Therefore, we have $S = \{s_1, s_2\}$ and $V = \{z_1, z_1', z_2, z_2', f\} \cup S$ where f represents a fruit. Let $L = \langle S, \pi \rangle$ be the associated stochastic F0L-system. π is determined by using Figure 6. Using Equation 2, we write

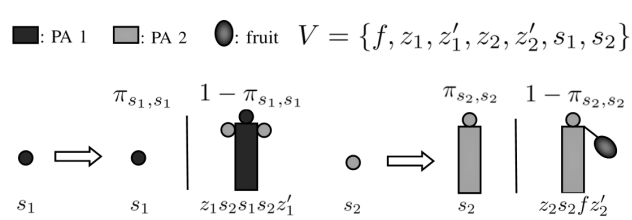

Figure 6. Stochastic F0L-system with fruits

the PGFs associated to $\mathcal{T}_L^n[s_i]$ on V for $i \in \{1,2\}$:

$$\Psi^n[s_i](f,\mathbf{z},\mathbf{s}) = \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_i}]} P_t f^{c(t,f)} \prod_{j=1}^{2}(z_j z_j')^{c(t,z_j)} \prod_{k=1}^{2}(s_k)^{c(t,s_k)} \tag{3}$$

with $\mathbf{z} = (z_1, z_1', z_2, z_2')$ and $\mathbf{s} = (s_1, s_2)$. Given that the elements of interest are fruits, the simplified PGFs are:

$$\Psi^n[s_i](f) = \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_i}]} P_t f^{c(t,f)} = \sum_{k\in\mathbb{N}} P_k^{n,i} f^k \quad , i \in \{1,2\} \tag{4}$$

$P_k^{n,i}$ is the probability that a plant of CA n with s_i as root carries k fruits.

Now, we need to write the structural property for PRTs with s_1 as root and s_2 as root. A root s_1 can be either a bud s_1 with a probability π_{s_1,s_1} or a metamer with one bud of PA 1 and 2 buds of PA 2 with a probability $1-\pi_{s_1,s_1}$. That means that a plant of CA $n+1$ with s_1 as root can either have the same structure as a plant of CA n with s_1 as root with a probability π_{s_1,s_1} or be a metamer connected to 3 plants of CA n (one with s_1 as root (= t_1) and 2 with s_2 as root (= t_2 and t_2')) with a probability $1-\pi_{s_1,s_1}$. Therefore, in the last case, we have $c(t,f) = c(t_1,f) + c(t_2,f) + c(t_2',f)$ and:

$$\Psi^{n+1}[s_1](f) = \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n+1}}[\mathrm{s_1}]} P_t f^{c(t,f)} = \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_1}]} \pi_{s_1,s_1} P_t f^{c(t,f)}$$
$$+ \sum_{(t_1,t_2,t_2')\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_1}]\times\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_2}]^2} (1-\pi_{s_1,s_1}) P_{t_1} P_{t_2} P_{t_2'} f^{c(t_1,f)+c(t_2,f)+c(t_2',f)}$$

Then:

$$\begin{aligned}\Psi^{n+1}[s_1](f) &= \pi_{s_1,s_1} \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_1}]} P_t f^{c(t,f)} \\ &+ (1-\pi_{s_1,s_1}) \left(\sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_1}]} P_t f^{c(t,f)}\right)\left(\sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_2}]} P_t f^{c(t,f)}\right)^2 \\ &= \pi_{s_1,s_1}\Psi^n[s_1](f) + (1-\pi_{s_1,s_1})\left(\Psi^n[s_1](f)\right)\left(\Psi^n[s_2](f)\right)^2\end{aligned} \tag{5}$$

A root s_2 can be either a bud s_2 with a probability π_{s_2,s_2} or a metamer with one bud of PA 2 and a fruit with a probability $1-\pi_{s_2,s_2}$. That means that a plant of CA $n+1$ with s_2 as root can either have the same structure as a plant of CA n with s_2 as root (= t_2) with a probability π_{s_2,s_2} or be a metamer connected to a plant of CA n with s_2 as root (= t_2') and a fruit (probability $1-\pi_{s_2,s_2}$). Therefore, in the last case, we have $c(t,f) = c(t_2',f)+1$ and:

$$\begin{aligned}\Psi^{n+1}[s_2](f) &= \sum_{t\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n+1}}[\mathrm{s_2}]} P_t f^{c(t,f)} \\ &= \sum_{t_2\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_2}]} \pi_{s_2,s_2} P_{t_2} f^{c(t_2,f)} + \sum_{t_2'\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_2}]} (1-\pi_{s_2,s_2}) P_{t_2'} f^{c(t_2',f)+1} \\ &= (\pi_{s_2,s_2} + (1-\pi_{s_2,s_2})f) \sum_{t_2\in\mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s_2}]} \pi_{s_2,s_2} P_{t_2} f^{c(t_2,f)} \\ &= (\pi_{s_2,s_2} + (1-\pi_{s_2,s_2})f)\,\Psi^n[s_2](f)\end{aligned}$$

Given that $\Psi^0[s_2](f) = 1$, then:

$$\Psi^n[s_2](f) = (\pi_{s_2,s_2} + (1-\pi_{s_2,s_2})f)^n \tag{6}$$

By identifying with Equation 4, we get $P_k^{n,2} = \binom{n}{k}\pi_{s_2,s_2}^{n-k}(1-\pi_{s_2,s_2})^k$ if $k \le n$ and 0 otherwise thanks to the binomial theorem.

By identifying the coefficients of the power series of Equation 5, we get:

$$P_k^{n+1,1} = \pi_{s_1,s_1} P_k^{n,1} + (1-\pi_{s_1,s_1}) \sum_{k_1+k_2+k_3=k} P_{k_1}^{n,1} P_{k_2}^{n,2} P_{k_3}^{n,2}$$

Initial conditions are given by $P_0^{0,1} = 1$ and $P_k^{0,1} = 0$ with $k \ge 1$. Thus, the distribution $\left\{P_k^{n,1}\right\}_{k\in\mathbb{N}}$ of the number of fruits for a plant of CA n can be determined numerically. The expected value E^n and the variance V^n are given by:

$$E^n = \sum_{k=0}^{+\infty} P_k^{n,1} k \quad \text{and} \quad V^n = \sum_{k=0}^{+\infty} P_k^{n,1}(k-E^n)^2$$

N.B.:

1) the sums in E^n and V^n contain only a finite number of terms. For $k > n$, we have $P_k^{n,1} = 0$.
2) Expected value and variance can be determined directly with Equations 5 and 6 by using classical properties of generating functions (see [14], [15]). By differentiating one and two times these equations with respect to f and by taking $f = 1$, we get a set of recursive equations containing E^n and V^n for $n \ge 0$.
3) This method is completely equivalent to that of [4]. By using Equation 3 for $n = 1$, we get:

   $$\Psi^1[s_1](f,\mathbf{z},\mathbf{s}) = \pi_{s_1,s_1} s_1 + (1-\pi_{s_1,s_1}) z_1 z_1' s_1 s_2^2$$

   Then, Equation 5 (the non simplified version) becomes:

   $$\Psi^{n+1}[s_1](f,\mathbf{z},\mathbf{s}) = \Psi^1[s_1](f,\mathbf{z},\Psi^n[s_1](f,\mathbf{z},\mathbf{s}),\Psi^n[s_2](f,\mathbf{z},\mathbf{s}))$$

   This composition is a classical property of branching processes (see [14]) and proves the equivalence of the two methods when counting permanent elements or structures. The next section illustrates an example showing the potential of the symbolic method and for which classical methods deriving from the generating functions of branching processes would fail.

5.2. Estimation of the age of a plant

In this section, we are interested in estimating the age of a plant by using botanical data. Let us assume that we have a plot of trees and that we know the proportions of the number of trees according to their age N, $\{P(N = n)\}_{n\in\{0,\ldots,N_m\}}$. N_m is the maximum CA. If we have no other information, we can simply give the expected value associated to the previous distribution: $E[N] = \sum_{k=0}^{N_m} nP(N = n)$. Therefore, the estimated age is the same for each tree.

Now, we want to be more precise and add botanical data to help the estimation. Let us assume that all the trees of the plot have grown according to the same stochastic F0L-system L and that we have a set of botanical data D (number of metamers, number of apexes, ...) for each tree. We can give an estimation of the CA taking into account all the data:

$$E[N|D] = \sum_{n=0}^{N_m} nP(N = n|D)$$

Bayes' theorem gives:

$$P(N = n|D) = \frac{P(\{N=n\} \cap D)}{P(D)}$$

Therefore:

$$E[N|D] = \sum_{n=0}^{N_m} n \frac{P(\{N = n\} \cap D)}{P(D)}$$

with $P(D) = \sum_{n=0}^{N_m} P(\{N = n\} \cap D)$. Then, to solve the problem we have to determine $P(\{N = n\} \cap D)$ by using Bayes' theorem again:

$$P(\{N = n\} \cap D) = P(D|N = n)P(N = n)$$

which is equivalent to find $P(D|N = n)$ for all $n \in \{0, \ldots, N_m\}$. To do so, we are going to use the symbolic method.

To make it clear, let us consider the case when D is the number of apexes. Let A be the random variable that gives the number of apexes in a tree generated by L. Then, we have to find $\{P(A = a|N = n)\}_{a \in \mathbb{N}}$. Let us compute this distribution in the case of a Leeuwenberg model with death probabilities. The stochastic automata is given by Figure 7. We only have one PA. Thus, $S = \{s\}$ and $V = \{z, z', d\} \cup S$ where d represents a dead bud. Let $L = \langle S, \pi \rangle$ be the associated stochastic F0L-system. π is determined by using Figure 7. Here we are interested in the distribution of the

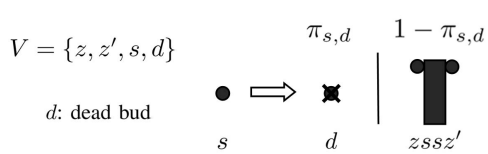

Figure 7. Leeuwenberg model with death probabilities

number of apexes. Each sequence of letters $zssz'$ and $zddz'$ corresponds to an apex (for example, the plant of Figure 2 has four apexes). Apexes can not be described by stochastic F0L-systems. For instance, at a given growth cycle, a word describing a plant can have patterns of type $t = zssz'$. Then, this kind of structures are apexes. However, it is possible that these structures evolve at the beginning of the next growth cycle and become $ztdz'$ for example. Thus, we do not have apexes anymore. Thus, these patterns depend on the structure of the plant and they can not be integrated in stochastic F0L-systems. In the sequel, every sequence "$zssz'$" and "$zddz'$" is designated by the letter a. The simplified PGF is:

$$\Psi^n[s](a) = \sum_{t \in \mathcal{T}_{\mathrm{L}}^{\mathrm{n}}[\mathrm{s}]} P_t a^{c(t,a)} = \sum_{k \in \mathbb{N}} P_k^n a^k \tag{7}$$

where $P_k^n = P(A = k|N = n)$. Hence, to determine the coefficients of $\Psi^n[s](a)$ is equivalent to solve the age problem.

As for structural properties, we have to break down the structure of a PRT in a way which highlights how apexes appear in the topology and how they are connected to substructures (see Figure 8). Let us suppose $n \geq 1$ and

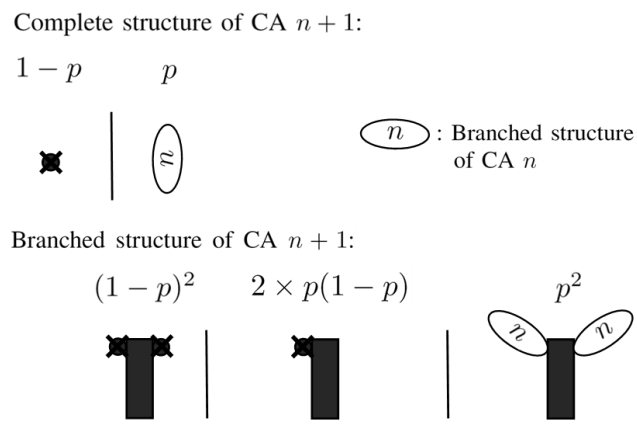

Figure 8. Structural property for Section 5.2

set $p = 1 - \pi_{s,d}$. The complete structure t of CA $n + 1$ corresponds always to one of these cases:

- a dead bud (probability $1 - p$). Thus, $c(t, a) = 0$.
- a branched structure t_1 of CA n. Thus, $c(t, a) = c(t_1, a)$.

Let $\mathcal{B}_{\mathrm{L}}^{\mathrm{n}}$ be the set of all possible branched structures of CA n generated by L. Then, we get:

$$\Psi^{n+1}[s](a) = (1 - p) + p \sum_{t_1 \in \mathcal{B}_{\mathrm{L}}^{\mathrm{n}}} P_{t_1} a^{c(t_1,a)}$$

Let $\Phi^n(a)$ be the simplified PGF associated to branched structures of CA n:

$$\Phi^n(a) = \sum_{t \in \mathcal{B}_{\mathrm{L}}^{\mathrm{n}}} P_t a^{c(t,a)} = \sum_{k \in \mathbb{N}} Q_k^n a^k \tag{8}$$

Then, we get:

$$\Psi^{n+1}[s](a) = (1 - p) + p\Phi^n(a) \tag{9}$$

A branched structure t of CA $n + 1$ corresponds always to one of these cases:

- a metamer with two dead buds (probability $(1 - p)^2$). Thus, $c(t, a) = 1$.
- a metamer with one bud and one branched structure t_1 of CA n (probability $p(1 - p)$ and two possible configurations). Thus, $c(t, a) = c(t_1, a)$.

- a metamer with two branched structures t_2 and t_3 of CA n (probability p^2). Thus, $c(t,a) = c(t_2,a) + c(t_3,a)$.

Then, we have:

$$\Phi^{n+1}(a) = (1-p)^2 * a + 2 * p(1-p) \sum_{t_1 \in \mathcal{B}_{\mathrm{L}}^{\mathrm{n}}} P_{t_1} a^{c(t_1,a)} + p^2 \sum_{(t_2,t_3)\in(\mathcal{B}_{\mathrm{L}}^{\mathrm{n}})^2} P_{t_2} P_{t_3} a^{c(t_2,a)+c(t_3,a)}$$

Then,

$$\Phi^{n+1}(a) = (1-p)^2 a + 2p(1-p)\Phi^n(a) + p^2 \left(\Phi^n(a)\right)^2 \quad (10)$$

By identifying the coefficients of the power series of Equation 10 and given that $\Phi^1(a) = a$, we get:

$$Q_k^{n+1} = (1-p)^2 \delta_1(k) + 2p(1-p) Q_k^n + p^2 \sum_{k_1+k_2=k} Q_{k_1}^n Q_{k_2}^n$$

where $\delta_j(k)$ is equal to 1 if $k = j$ and 0 otherwise. Finally, by identifying the coefficients of the power series of Equation 9, we have:

$$P_k^{n+1} = (1-p)\delta_0(k) + pQ_k$$

For $n = 0$, we have $P_k^0 = \delta_0(k)$. Now, we have determined $\{P(A = a|N = n)\}_{a\in\mathbb{N}}$ and we can estimate more precisely the age of a plant by counting its apexes.

6. Conclusion

We have set a new mathematical framework to describe both plant topology and development. On the one hand, we used a combinatorial approach to represent plants with plane rooted trees and we gave an equivalent representation using Dyck words. This approach enables an accurate description of plant topology. On the other hand, we adapted the formalism relying on stochastic F0L-systems to integrate plant development. The combination of these two components allows the use of effective combinatorial tools such as the symbolic method. As a matter of fact, we are now able to compute the distribution and the moments of all kind of organs or structures in a plant. In that sense, this method is a generalization of that developped in [4]. We are now able to study particular structures in plants such as apexes. Thus, it opens new kinds of applications. In the sequel, it might be interesting to use this approach for statistical purposes. As a matter of fact, knowing the distribution of all kind of elements in a plant will allow accurate estimations of parameters in a growth model.

References

[1] P. de Reffye, M. Goursat, J. Quadrat, and B. Hu, "The Dynamic Equations of the Tree Morphogenesis Greenlab Model," INRIA, Tech. Rep. 4877, 2003.

[2] P.-H. Cournède, M.-Z. Kang, A. Mathieu, J.-F. Barczi, H.-P. Yan, B.-G. Hu, and P. de Reffye, "Structural Factorization of Plants to Compute their Functional and Architectural Growth," *Simulation*, vol. 82, no. 7, pp. 427–438, 2006.

[3] M.-Z. Kang, P.-H. Cournède, J.-P. Quadrat, and P. de Reffye, "A stochastic language for plant topology," in *Plant growth Modeling, simulation, visualization and their Applications.*, T. Fourcaud and X. Zhang, Eds. IEEE Computer Society (Los Alamitos, California), 2007.

[4] C. Loi and P.-H. Cournède, "Generating functions of stochastic L-systems and application to models of plant development," *proceedings of Discrete Mathematics and Theoretical Computer Science*, 2008.

[5] P. Kruszewski and S. Whitesides, "A General Random Combinatorial Model of Botanical Trees." *Journal of Theoretical Biology*, vol. 191, pp. 221–236, 1998.

[6] D. Barthélémy and Y. Caraglio, "Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny," *Annals of Botany*, vol. 99, no. 3, pp. 375–407, 2007.

[7] P. Flajolet and R. Sedgewick, *Analytic Combinatorics*. Cambridge University Press, 2009.

[8] J. Riordan, *An Introduction to Combinatorial Analysis*. Courier Dover Publications, 2002.

[9] D. Knuth, *The Art of Computer Programming*. Addison-Wesley Professional, 1997 (3rd edition), vol. 1.

[10] P. Prusinkiewicz and A. Lindenmayer, *The Algorithmic Beauty of Plants*. Springer-Verlag, New-York, 1990.

[11] J. Françon, "Sur la modélisation informatique de l'architecture et du développement des végétaux." in *2ème Colloque International: L'Arbre. Institut de Botanique, Montpellier, France*, 1990.

[12] W. Kurth, *Growth grammar interpreter GROGRA 2.4: A software tool for the 3-dimensional interpretation of stochastic, sensitive growth grammars in the context of plant modelling. Introduction and Reference Manual*. Berichte des Forschungszentrums Waldökosysteme der Universität Göttingen, Ser. B, Vol. 38, 1994.

[13] M.-Z. Kang, P.-H. Cournède, P. de Reffye, D. Auclair, and B.-G. Hu, "Analytical study of a stochastic plant growth model: application to the greenlab model," *Mathematics and Computers in Simulation*, vol. 78, no. 1, pp. 57–75, 2008.

[14] T. Harris, *The theory of branching processes*. Springer, Berlin, 1963.

[15] C. Mode, *Multitype branching processes: Theory and applications*. American Elsevier Publishing Co. Inc, New York, 1971.