A Symbolic Method to Analyse
Patterns in Plant Structure

C. L O I[1], P.-H. C O U R N È D E[1] and J. F R A N Ç O N[2]

[1]Ecole Centrale Paris, Laboratory of Applied Mathematics and Systems, Châtenay Malabry, France
(e-mail: `paul-henry.cournede@ecp.fr`)

[2]University of Strasbourg, ICube Laboratory, Team 'Informatique Géométrique et Graphique', France

Formal grammars such as L-systems have long been used to describe plant growth dynamics. In this article, they are used for a new purpose. The aim is to build a symbolic method that enables the computation of the stochastic distribution associated with the number of complex structures in plants whose organogenesis is driven by a multitype branching process. For that purpose, a new combinatorial framework is set in which plant structure is coded by a Dyck word. Moreover, organogenesis is represented by stochastic F0L-systems. In doing so, the problem is equivalent to determining the distribution of patterns in random words generated by a stochastic F0L-system. This method finds interesting applications in the parameter identification of stochastic models of plant development.

2010 *Mathematics subject classification*: Primary 05A
Secondary 92B, 60J

1. Introduction

In botany, organogenesis is the process during which the buds of a plant produce new organs. Formal grammars have long been used to describe plant organogenesis. In particular, the parallel rewriting grammar introduced by [22] (called the L-system) is well adapted to modelling the development of branching patterns, and its algorithmic power has been widely exploited since [36, 32]. It has thus provided efficient algorithms and subsequently software languages for plant simulation (see [32, 21, 16, 17]). For structures as complex as trees, bud production rules are influenced by many factors and are usually modelled by stochastic processes. In that case, organogenesis can be represented by stochastic 0L-systems [32, 5, 25, 23]. This type of grammar gives interesting results from simulation and graphic-theoretic perspectives by increasing the realistic aspect of virtual plants (see [32, 27, 13]). However, in the context of plant growth modelling, the full mathematical potential of L-systems [35, 8] has not been fully exploited.

In this article, stochastic 0L-systems are used for a new purpose. The objective is to write a method allowing the computation of the distribution associated with the number

of complex structures in plants whose organogenesis is driven by a multitype Galton–Watson branching process. This method relies on a symbolic approach derived from combinatorics and the analysis of algorithms (see [9]). Plants are seen as combinatorial structures. In this new combinatorial framework, plant topology is described by a Dyck word [19], which is an original contribution, since other representations such as bracketed strings [32] are generally used. The development of the structure of the plant is given by a set of stochastic rules contained in a stochastic F0L-system. Therefore, this framework allows the use of powerful methods of combinatorics such as the symbolic method [9]. Including a symbolic approach in a dynamic branching structure gives a more complete description of the system. It enables the computation of the distribution of patterns in a sequence of words whose dynamic evolution is driven by a branching process. Such results find applications in the comparison and parameter identification of stochastic models.

Some basic concepts of botany and the main features of stochastic organogenesis models driven by a multitype branching process are first recalled in Section 2. Then, a new combinatorial framework is set up in Section 3. Plants are seen as labelled ordered rooted trees. This allows the description of their structures by Dyck words. The development rules of the organogenesis model are represented by a stochastic F0L-system. Section 4 proposes a symbolic method adapted to plant growth models. In the last section, we show how to use the symbolic method for the parameter identification of the stochastic processes during plant development.

2. Stochastic modelling of organogenesis with a multitype branching process

Models of plant development (or organogenesis) describe the dynamic creation of organs (internodes, buds, leaves, flowers or fruits) and how they arrange themselves to form plant structure. When the smallest scale of interest is that of organs (and not cells), discrete models are generally used to simulate plant structural development.

2.1. Modelling of plant structure

In this article, only the above-ground parts of plants are considered. As explained for example in [3], organogenesis results from the functioning of undifferentiated cells constituting the apical meristem and located at the tip of axes. When in the active phase, this meristem forms buds that will develop into agglomerates of organs composed of one or several *phytomers*. A phytomer is a botanical entity chosen as the elementary unit used to represent the plant architecture. It is composed of an internode and one or more leaves (see Figure 1). Depending on the type of plant, the internode may also bear flowers and fruits. When a phytomer (or group of phytomers) is built, the apical meristem forms a new terminal bud and the internode may often bear lateral buds. These buds may potentially generate new phytomers later on.

Plant development can be discretized in time. In plants with rhythmic growth, a terminal bud produces a set of subsequent phytomers, called a *growth unit* (or *shoot*), and then rests for a while. The time between the appearance of new growth units (*i.e.*, the time step of the discrete model) defines the *growth cycle* [6]. Consecutive growth units form axes.
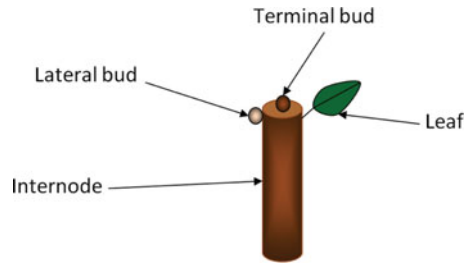
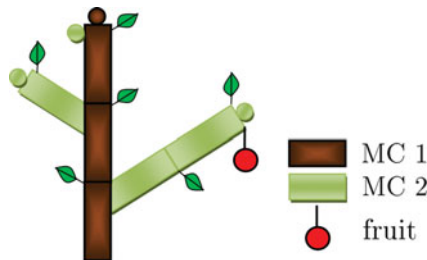*Figure 1.* (Colour online) Schematic representation of a phytomer

.

Figure 2. (Colour online) Example of a plant with two morphological categories (MC). The fruit has no MC attached.

Concerning the architecture of the plant, axes and architectural units can be divided into different categories depending on their morphological parameters. These categories will be called *morphological categories* (MC) below. Other names can be found in the literature. (For example, [2] introduced the concept of physiological age to represent the different types of axes resulting from the meristematic differentiation in a plant. The concept of MC is more general and is chosen to allow for cases that do not correspond to the strict botanical concept of physiological age.) By convention, the terminal bud of an axis is thus characterized by the MC of the phytomer that bears it (Figure 2 gives an example of a plant with two MCs).

Here, an organ will be characterized by its type (*e.g.*, internode, bud, fruit) and by its botanical characteristics (*e.g.*, MC, chronological age).

2.2. Stochastic organogenesis

Plant development can be discretized in time. The time between the appearances of new growth units (or shoots) (*i.e.*, the time step of the discrete model) defines the *growth cycle* (GC). For example, most temperate trees grow rhythmically, new shoots appearing at spring. If we do not consider polycyclism and neoformation, the GC corresponds to one year. The *chronological age* (CA) of a plant (or an organ) is defined as the number of GCs for which it has existed.

The structure of a plant changes from one GC to another. For example, a bud may create new organs or the botanical characteristics of an organ may change (such as its CA). A set of rules called *development rules* defines the way organs evolve from one GC
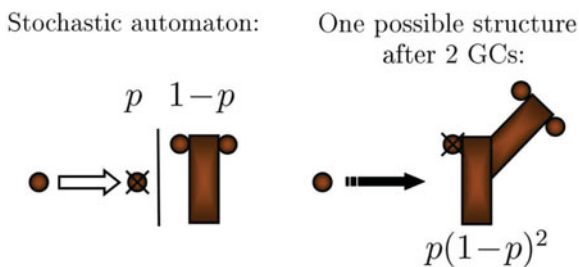
Figure 3.   (Colour online) Example of a stochastic automaton for an organogenesis model.

to another. In a stochastic organogenesis model, an organ may have several possible evolutions. In that case, a probability of occurrence is associated with each of these. As explained in [15], the probabilities involved in the organogenesis model are the result of botanical phenomena (*e.g.*, bud survival, bud dormancy, differentiation).

In this paper we consider a class of organogenesis models satisfying the following hypotheses.

(1) At each GC, the organs behave independently: the evolution of an organ is not influenced by the other organs of the plant.
(2) The probabilities of evolution associated with an organ depend only on its type and its characteristics.

Under these hypotheses, a plant structure of CA N is built recursively by using the following algorithmic procedure.

- Initialization: the structure of the plant at GC 0 is given by a seed (*i.e.*, a bud of MC 1).
- For all $n = 0 \cdots N - 1$, the plant structure at GC $n + 1$ is built from the plant structure at GC n by replacing all the organs randomly by one of their possible evolutions according to their probabilities of occurrence.

The underlying stochastic process associated with this class of organogenesis models is a multitype Galton–Watson branching process (see [28] and [1]). Galton–Watson branching processes are frequently used to describe development processes in biology from the development of cell populations (see [4] for angiogenesis or [14] for plant pathogens). Likewise, organs created to build plant structure can be seen as individuals in a population process. Therefore, the previous algorithmic procedure creates a Markovian sequence of random plants indexed by the GCs. The development rules associated with the organogenesis model are usually represented by a set of stochastic automata (see [36] and [32] and Figure 3).

Figure 3 shows an example of an organogenesis model with one MC; p represents the death probability for a bud. In this example, a bud has probability p of dying and probability $1 - p$ of creating one internode with two lateral buds. Note that there is no stochastic automaton concerning the possible evolutions of an internode. In that case, it means that internodes do not evolve from one GC to another (they stay in the same state
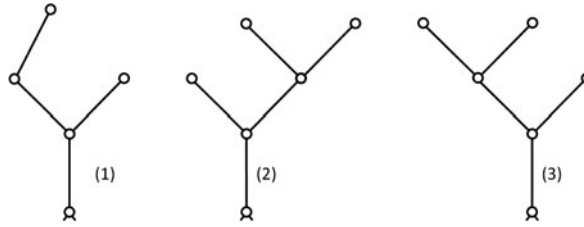
Figure 4. Examples of ordered rooted trees.

with probability equal to 1). The plant on the right is one possible structure occurrence after two GCs. The associated probability of occurrence is $p(1 - p)^2$.

Here, $\mathcal{S}$ will denote a stochastic organogenesis model driven by a multitype branching process. A 'plant' refers to a branched structure generated from a seed (*i.e.*, a bud of MC 1) according to the development rules of $\mathcal{S}$. In the following section, a combinatorial framework is set up to describe plant architecture and its development. The structure of a plant is given by a Dyck word in a bijective way. Since the organogenesis is driven by a multitype Galton–Watson branching process, the development rules can be represented by a stochastic 0L-system (see [23] for more details).

3. Plants as combinatorial structures

3.1. Some combinatorial concepts

We recall some basic definitions and properties of combinatorics (see [9] and [34] for more details).

Definition 3.1 (ordered rooted tree (ORT)). A rooted tree is a connected and acyclic graph with a node specifically distinguished called root. A plane tree is defined as a tree in which subtrees coming from a common node are ordered and represented from left to right.

Figure 4 gives examples of an ORT. The trees (2) and (3) are equivalent as rooted trees, but they become distinct objects when regarded as ORTs.

ORTs can be coded in several ways, one of the most classical being the Dyck word coding [19]. The latter relies on a tree traversal using the prefix order (see [9, p. 74]).

Definition 3.2 (prefix order). An ORT is traversed according to the prefix order if it is traversed starting from the root, proceeding depth-first and left-to-right, and backtracking upwards once a subtree has been completely traversed.

Figure 5 gives an example of preorder traversal.

Definition 3.3 (Dyck word coding). Every ORT is described by a Dyck word on the alphabet $V_1 = \{z, z'\}$ as follows.

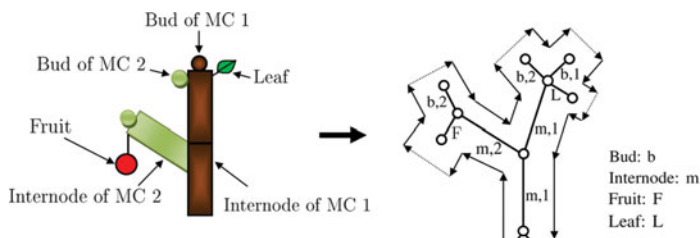• The tree is traversed according to the prefix order.

Figure 5. (Colour online) Correspondence between plants and labelled ordered rooted trees. For the sake of clarity, only the labels of edges are represented.

- An edge visited from the parent node to the child node is represented by the letter z.
- An edge visited from the child node to the parent node is represented by the letter z'.

In the case of labelled trees (trees with characteristics attached to nodes and edges), a new code deriving from Dyck word coding can be used.

Definition 3.4 (extended Dyck word coding). Let $L = \{l_1, \ldots, l_M\}$ be a set of labels. Every labelled ORT with labels in L is described by an extended Dyck word on the alphabet $V_2 = \{z_{l_1}, z'_{l_1}, \ldots, z_{l_M}, z'_{l_M}\}$ as follows.

- The tree is traversed according to the prefix order.
- An edge with a label l_i and visited from the parent node to the child node is represented by the letter z_{l_i}.
- An edge with a label l_i and visited from the child node to the parent node is represented by the letter z'_{l_i}.

3.2. Coding a plant structure with a Dyck word

Every plant generated by the organogenesis model $\mathcal{S}$ can be represented by a labelled ORT.

Definition 3.5 (labelled ORT associated with a plant). Let P be a plant generated by the organogenesis model $\mathcal{S}$. Let us build a labelled ORT from the plant P as follows.

- Every organ is represented by an edge and ends with a node at each extremity.
- Two connected organs are represented by two edges having a common node.
- The root is the node below the edge representing the internode at the basis of the plant.
- The label of an edge is given by the botanical information of the associated organ: its nature (*e.g.*, bud, leaf, internode) and its characteristics (*e.g.*, MC, CA).
- The label of a node (different from the root) is the label of the edge below it. The root has the label of the edge above it.

Such an ORT is called the labelled ORT associated with the plant P.

Note. The labelled ORT defined in Definition 3.5 is unique for a given plant structure and the converse is true.

Since a plant can be represented by a labelled ORT, it has an associated extended Dyck word. Let $\mathcal{O}^{\mathcal{S}}$ be the minimal set of letters coding for the type of organs needed to describe plant structures generated by $\mathcal{S}$. Generally, the letter b codes for a bud, m for an internode, L for a leaf and F for a fruit. In the same way, let $\mathcal{C}^{\mathcal{S}}$ be the minimal set of all possible characteristics associated with $\mathcal{S}$. In Figure 5, the extended Dyck word associated with the plant (*i.e.*, its ORT) is thus $w = z_{m,1}z_{m,2}z_F z'_F z_{b,2}z'_{b,2}z'_{m,2}z_{m,1}z_{b,2}z'_{b,2}z_{b,1}z'_{b,1}z_L\ z'_L z'_{m,1}z'_{m,1}$ where the first letter of a label represents the type of an organ and the second its MC. Note that no MC is associated with leaves and fruits.

For the sake of clarity, some coding conventions will be used.

- Every organ of nature $o \in \mathcal{O}^{\mathcal{S}}$ having the characteristics $c \in \mathcal{C}^{\mathcal{S}}$ is represented by the symbols o_c and o'_c (instead of $z_{o,c}$ and $z'_{o,c}$).
- For some organs (*e.g.*, buds, leaves, fruits), it is not necessary to represent the visit from the child node to the parent node of the associated edge. In fact, a visit from the parent node to the child node is always immediately followed by a visit from the child node to the parent node. Thus, only the symbol representing the visit from the parent node to the child node is used to describe that type of organ.

By taking into account the previous conventions of coding, the plant of Figure 5 is thus coded by $w = m_1 m_2 F b_2 m'_2 m_1 b_2 b_1 L m'_1 m'_1$.

The set of all labelled ORTs generated by organogenesis model $\mathcal{S}$ is denoted by $\mathcal{T}^{\mathcal{S}}$. Let $V^{\mathcal{S}} = \{o_c, o'_c\}_{o \in \mathcal{O}^{\mathcal{S}}, c \in \mathcal{C}^{\mathcal{S}}}$ be the minimal alphabet containing all the letters needed to describe plant structures generated by $\mathcal{S}$, let $W_{V^{\mathcal{S}}}$ be the set of all words built on $V^{\mathcal{S}}$ and let $DW_{V^{\mathcal{S}}}$ be the set of all extended Dyck words generated by $\mathcal{S}$ (with the previous conventions of notation). Let $D_{V^{\mathcal{S}}} : \mathcal{T}^{\mathcal{S}} \to DW_{V^{\mathcal{S}}}$ be the map which associates for each labelled ORT t its corresponding extended Dyck word $D_{V^{\mathcal{S}}}(t)$. Then, $D_{V^{\mathcal{S}}}$ is a bijection from $\mathcal{T}^{\mathcal{S}}$ to $DW_{V^{\mathcal{S}}}$ [19].

Remark. In the plant modelling community, other representations for trees exist, particularly bracketed strings, with brackets to delimit branching structures [32]. The plant in Figure 5 could be coded by $w = m_1[m_2[F]b_2]m_1[L][b_2]b_1$. The symbolic method described in the following sections to compute the distributions of patterns can similarly be applied with such a representation. The introduction of this representation of plants as Dyck words proves interesting since this new representation is often more synthetic for very branched structures or structures without clear hierarchy in branching orders (no main trunk, for example, such as for plants of the Leeuwenberg-type growth model [11]). Such characteristic is crucial for combinatorial purposes.

3.3. Growth dynamics and L-systems

In Section 3.2 it was proved that the structure of a plant at a given growth cycle can be represented by an extended Dyck word. We are now interested in the development of the structure with its chronological age starting from a seed (or a bud). As for stochastic

organogenesis models driven by multitype branching processes, this evolution is given by a Markovian sequence of random plants indexed by the growth cycles (see Section 2.2). Thus, it can be described by a sequence $(t^n)_{n \in \mathbb{N}}$ of $\mathcal{T}^\mathcal{S}$ (or the corresponding sequence $(D_V(t^n))_{n \in \mathbb{N}}$ of $DW_{V\mathcal{S}}$). To complete the combinatorial framework, we need to define tools which describe the stochastic organogenesis model when the structure is coded by an extended Dyck word (*i.e.*, tools which enable the building of $D_V(t^{n+1})$ from $D_V(t^n)$).

Stochastic 0L-systems are well suited to achieving this goal (see [32] and [20]). They are generative parallel rewriting grammars whose production rules are associated with a set of probability distributions.

Definition 3.6 (stochastic 0L-system). Let V be an alphabet and let W_V be the set of all words built on V. A stochastic 0L-system is a construct $L = \langle \omega_a, \pi \rangle$ where:

- $\omega_a \in W_V$ is called the axiom,
- π is a transition matrix from V to W_V (*i.e.*, for all $(u,v) \in V \times W_V$, $0 \leqslant \pi_{u,v} \leqslant 1$ and $\sum_{w \in W_V} \pi_{u,w} = 1$) with a finite number of non-zero components.

A stochastic 0L-system $L = \langle \omega_a, \pi \rangle$ generates a random sequence of words $(w^n)_{n \in \mathbb{N}}$ built on the alphabet V. By definition, the axiom is the word which initiates the sequence generated by L. Then, $w^0 = \omega_a$. We get w^{n+1} by replacing randomly every letter x of w^n by a word y with probability $\pi_{x,y}$ (note that the evolution of a letter is independent from the evolution of the other letters). By doing so, we create a Markov chain on W_V: $(w^n)_{n \in \mathbb{N}}$. We can now define a more general class of 0L-systems called the stochastic F0L-system, extending the classical definition of the F0L-system [35, p. 89] to the stochastic case.

Definition 3.7 (stochastic F0L-system). Let V be an alphabet and let W_V be the set of all words built on V. A stochastic F0L-system is a construct $L = \langle A, \pi \rangle$ where:

- A is a non empty subset of W_V (called the set of axioms of L),
- for every $\omega_a \in A$, $L[\omega_a] = \langle \omega_a, \pi \rangle$ is a stochastic 0L-system (called the component system of L).

The stochastic organogenesis model $\mathcal{S}$ can be described by a stochastic F0L-system $L = \langle W_{V\mathcal{S}}, \pi \rangle$. The development rules of $\mathcal{S}$ are described by a set of stochastic automata (see Section 2.2). This set is used to fill the transition matrix π. Let us take the example of Figure 6. From the stochastic automata, we deduce $\pi_{b_1,b_1} = p$, $\pi_{b_1,m_1b_2b_1m'_1} = 1 - p$ and $\pi_{b_2,m_2b_2m'_2} = 1$. Note that the development rules concerning the internodes of MC 1 and MC 2 are not specified. In that case, it means that they stay in the same state from one GC to another. Therefore, $\pi_{m_1,m_1} = 1$ and $\pi_{m_2,m_2} = 1$. All the other components of π are equal to 0. The stochastic F0L-system $L = \langle W_{V\mathcal{S}}, \pi \rangle$ generates a Markov chain on $DW_{V\mathcal{S}}$, $(w^n)_{n \in \mathbb{N}}$, coding for plant structures. The associated Markov kernel Π (called the Markov kernel associated with L) can be built very easily from π (see [23] for more details). For all $(u,v) \in W_{V\mathcal{S}} \times W_{V\mathcal{S}}$, $(\Pi^n)_{u,v}$ is the probability of getting the word v by using the stochastic 0L-system $L[u]$ after n steps; w^0 codes for the seed (*i.e.*, a bud of MC 1). Since the development rules of $\mathcal{S}$ are contained in π, w^{n+1} is built randomly from w^n by using
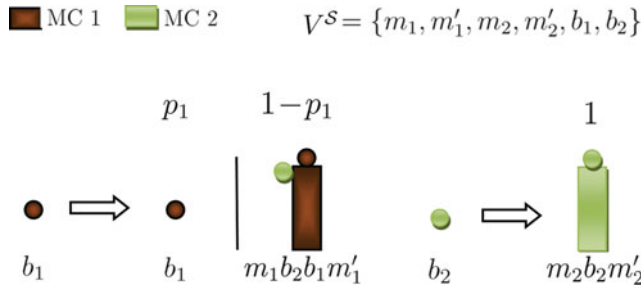
Figure 6. (Colour online) Stochastic automata and the associated stochastic F0L-system.

the same rules as those described by $\mathcal{S}$. Therefore, studying the development of plant structures generated by $\mathcal{S}$ is completely equivalent to studying the sequence $(w^n)_{n\in\mathbb{N}}$ from a combinatorial perspective.

Here $L = \langle W_{V^{\mathcal{S}}}, \pi \rangle$ will denote a stochastic F0L-system associated with the stochastic organogenesis model $\mathcal{S}$. When there is no confusion as to the model used, the letter $\mathcal{S}$ will not be specified in the corresponding sets. In that case, $L = \langle W_V, \pi \rangle$. Π will denote the Markov kernel associated with L. For all $s \in W_V$, $DW_V^{\pi,n}[s]$ ($\subset DW_V$) denotes the set of all possible extended Dyck words generated by the component system $L[s]$ after n steps. Let $\mathcal{T}^{\pi,n}[s] = D_V^{-1}(DW_V^{\pi,n}[s])$ be the set of all labelled ORTs ($\mathcal{T}^{\pi,n}[s] \subset \mathcal{T}$) generated by $L[s]$ after n steps.

4. The symbolic method

The symbolic method is a very effective method for analysing combinatorial structures and, as a consequence, plays an important role in analytic combinatorics (see [9] for more details). The complete theoretical framework associated with this section can be found in [24]. As far as plants are concerned, it enables us to compute not only the distribution of the number of organs (of any type) but also of specific structures in the plant architecture.

4.1. Combinatorial class and generating function
In this section, basic tools of combinatorics are recalled (see [9] for more details).

Definition 4.1 (combinatorial class). A combinatorial class, or simply a class, is a finite or denumerable set on which a size function is defined, satisfying the following conditions:

- the size of an element is a non-negative integer,
- the number of elements of any given size is finite.

For example, for all $n \in \mathbb{N}$, $DW_V^{\pi,n}[s]$ is a combinatorial class. Many size functions can be defined (*e.g.*, counting the number of letters coding for internodes).

Definition 4.2 (stochastic combinatorial class). A stochastic combinatorial class is a set

$$SC = \{(t, p_t), t \in C\}$$

such that:

- C is a combinatorial class,
- for all $t \in C$, $0 \leqslant p_t \leqslant 1$,
- $\sum_{t \in C} p_t = 1$.

The set $SDW_V^{\pi,n}[s] = \{(w, (\Pi^n)_{s,w}), w \in DW_V^{\pi,n}[s]\}$ is a stochastic combinatorial class.

Definition 4.3 (generating function (GF) associated with a size function in a stochastic combinatorial class). Let C be a combinatorial class and let $SC = \{(t, p_t), t \in C\}$ be a stochastic combinatorial class. Let m be a size function in C. The generating function ψ associated with m in SC is a mapping from $[0,1]$ to $[0,1]$ defined as follows:

$$\text{for all } z \in [0,1], \quad \psi(z) = \sum_{t \in C} p_t \; z^{m(t)}.$$

GFs are very useful for analysing a stochastic combinatorial class SC from a probabilistic perspective. Suppose we are interested in getting the distribution of a particular element c in SC. Let m be the size function ($SC \to \mathbb{N}$) such that, for all $t \in SC$, $m(t)$ gives the number of elements c in t. By reordering the terms of the GF, we get the following power series:

$$\text{for all } z \in [0,1], \quad \psi(z) = \sum_{t \in C} p_t \; z^{m(t)} = \sum_{k \in \mathbb{N}} p(k) z^k.$$

Here $p(k)$ is the probability of getting k elements c in SC. Therefore, by determining the coefficients of the GF associated with the size function m in SC, we get the distribution of the number of c in SC.

4.2. Description of the method

Suppose we are interested in computing the distribution of a particular structure (*e.g.*, a specific sequence of phytomers or a particular element in the plant) in a plant of CA n generated by an organogenesis model $\mathcal{S}$. To solve this issue in the deterministic case, there is the possibility of simulating the organogenesis model and then exploring the graph structure obtained, as proposed by the GroIMP software [18]. In the stochastic case, the combinatorial complexity of the organogenesis model makes the Monte Carlo simulation approach difficult, since the number of samples necessary to approximate the distribution of patterns may be prohibitively large, specifically since the exploration of the graph structure is also costly. The theoretical approach that we propose here is thus an interesting alternative.

The particular structure of interest can be coded on the alphabet V by a word u. Therefore, this problem is equivalent to determining the distribution of elements u in the stochastic combinatorial class $SDW_V^{\pi,n}[s]$, for the size function m chosen as the function counting the number of elements u, as detailed below. The characteristic word u of interest will be called a *pattern*.

Definition 4.4 (counting function). Let V be an alphabet and let W_V be the set of all words built on V. The counting function c is a map from $W_V \times W_V$ to $\mathbb{N}$ such that, for all $(w, u) \in W_V \times W_V$, $c(w, u)$ gives the number of words u in the word w.

Note that this definition implies that two overlapping words are counted as two different elements. In order to exclude overlapping, some additional rules have to be included (*e.g.*, a priority rule based on the order of appearance) that induce some technical adaptations below.

For all $u \in W_V$ and $n \in \mathbb{N}$, the mapping $w \mapsto c(w, u)$, from $DW_V^{\pi,n}[s]$ to $\mathbb{N}$, is a size function for the combinatorial class $DW_V^{\pi,n}[s]$. Therefore, to get the distribution of a pattern u in $SDW_V^{\pi,n}[s]$, we need to compute the GF associated with the size function $w \mapsto c(w, u)$ in $SDW_V^{\pi,n}[s]$. Henceforth we will call the GF associated with a pattern u in $SDW_V^{\pi,n}[s]$ the GF denoted by $\psi^n[s](z)$ and given by:

$$\text{for all } z \in [0,1], \quad \psi^n[s](z) = \sum_{w \in DW_V^{\pi,n}[s]} (\Pi^n)_{s,w} z^{c(w,u)} = \sum_{k \in \mathbb{N}} P^{n,s}(k) z^k,$$

where $P^{n,s}(k)$ is the probability of getting k patterns u in a plant structure generated by $L[s] = \langle s, \pi \rangle$ after n steps. However, this GF is rarely determined directly. Usually, we get it from functional equations which, most of the time, are recurrence relations between $\psi^n[s]$ and $\psi^{n+1}[s]$. To obtain these equations, we use a symbolic approach as developed in [9].

Let C be a combinatorial class and let $SC = \{(t, p_t), t \in C\}$ be a stochastic combinatorial class. Suppose we are interested in a particular size function m taking its argument in C. Thus, we want to determine the GF ψ associated with m in SC. The idea of the symbolic method is to build an equation which decomposes SC into smaller classes, either of the same type or of simpler types (we will refer to this equation as the set equation). Then, it is transformed into an equation composed of the GFs of the previous combinatorial classes. If possible, we solve the transformed equation directly and we get ψ. When the structure of the class SC is too complex, we extract recurrence relations satisfied by the coefficients of ψ from the transformed equation.

When dealing with plants, for a given $n \in \mathbb{N}$, we have $C = DW_V^{\pi,n}[s]$ and $SC = SDW_V^{\pi,n}[s]$, where s is an extended Dyck word coding for a seed. The idea is to decompose $DW_V^{\pi,n}[s]$ into a combination of sets of type $DW_V^{\pi,k}[s']$ with $k < n$ and $s' \in W_V$. By doing so, we get a set equation which can be transformed into an equation composed of GFs.

The symbolic method can be decomposed into the following steps.

(1) Identify the structure of the plant and all stochastic behaviours.
(2) Deduce the associated stochastic automata.
(3) Find the appropriate alphabet V to describe the plant and write the stochastic F0L-system $L = \langle W_V, \pi \rangle$.
(4) Identify the pattern $u \in W_V$ whose distribution you want to compute.
(5) Write the structural property associated with each $DW_V^{\pi,n}[s]$ for $n \in \mathbb{N}$ where s is an extended Dyck word coding for a seed and try to decompose it into a combination of sets of type $DW_V^{\pi,k}[s']$ with $k < n$ and $s' \in W_V$.
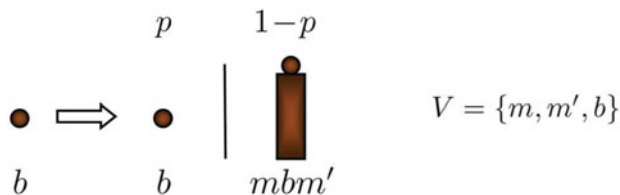
Figure 7. (Colour online) Plant with one morphological category and dormancy probabilities.

(6) Write the transformed equation satisfied by the GF $\psi^n[s]$ associated with u in $SDW_V^{\pi,n}[s]$ for each $n \in \mathbb{N}$.

(7) Either solve the set of transformed equations directly or find a recurrence relation between $\psi^n[s]$ and $\psi^{n+1}[s]$ with $n \in \mathbb{N}$.

(8) Extract from these equations the coefficients of the GFs.

4.3. Examples

4.3.1. Example with simple elements. We want to compute the distribution associated with the number of internodes in the following plant growth model: a plant with only one MC and a dormancy probability p. The behaviour of a bud is characterized by Figure 7. The alphabet is simply $V = \{m, m', b\}$, where m codes for an internode and b for a bud. The transition matrix of the associated stochastic F0L-system $L = \langle W_V, \pi \rangle$ can be easily deduced from Figure 7. We have $\pi_{b,b} = p$, $\pi_{b,bmb'} = 1 - p$, $\pi_{m,m} = 1$ and $\pi_{m',m'} = 1$. All the other components of π are equal to zero. The pattern of interest is m. From Figure 7, we deduce that a word $w \in DW_V^{\pi,n+1}[b]$ can be:

- either a word $v \in DW_V^{\pi,n}[b]$ with probability p,
- or a word of the form mvm' with $v \in DW_V^{\pi,n}[b]$ with probability $1 - p$.

Let $\psi^n[b]$ be the GF associated with m in $DW_V^{\pi,n}[b]$:

$$\psi^n[b](z) = \sum_{w \in DW_V^{\pi,n}[b]} (\Pi^n)_{b,w} \, z^{c(w,m)} = \sum_{k \in \mathbb{N}} P^{n,b}(k) z^k$$

where $P^{n,b}(k)$ is the probability of getting k internodes in a plant structure generated by the organogenesis model described in Figure 7 after n GCs. Then we have

$$\psi^{n+1}[b](z) = \sum_{v \in DW_V^{\pi,n}[b]} p(\Pi^n)_{b,v} \, z^{c(v,m)} + \sum_{v \in DW_V^{\pi,n}[b]} (1-p)(\Pi^n)_{b,v} \, z^{c(mvm',m)}.$$

Given that $c(mvm', m) = c(v, m) + 1$, we have

$$\psi^{n+1}[b](z) = p \sum_{v \in DW_V^{\pi,n}[b]} (\Pi^n)_{b,v} \, z^{c(v,m)} + (1-p)z \sum_{v \in DW_V^{\pi,n}[b]} (\Pi^n)_{b,v} \, z^{c(v,m)},$$

and then

$$\psi^{n+1}[b](z) = p\psi^n[b](z) + (1-p)z\psi^n[b](z) = (p + (1-p)z)\psi^n[b](z). \tag{4.1}$$
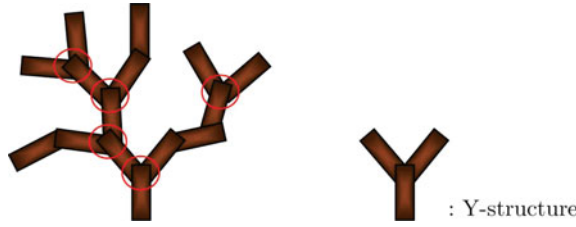
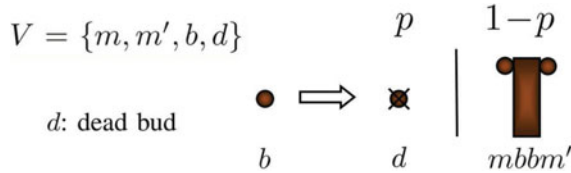Figure 8. (Colour online) Illustration of Y-structures in a plant.

Figure 9. (Colour online) Leeuwenberg-type growth model with death probability.

Given that $\psi^0[b] = 1$, the solution of the previous geometric progression is then

$$\psi^n[b](z) = (p + (1 - p)z)^n.$$

We deduce $P^{n,b}(k) = \binom{n}{k} p^{n-k}(1 - p)^k$ if $0 \leqslant k \leqslant n$, and 0 otherwise.

Note. As detailed in [23], the underlying stochastic process in this section is that of a Galton–Watson multitype branching process (see [28] and [1]). In fact, let B_n and M_n be two random variables on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ (where $\mathbf{P}$ is a probability measure) such that B_n and M_n, respectively, give the number of letters b and m in a word generated randomly by either the component system $L[b]$ or the component system $L[m]$ after n steps. Then, the sequence of random vectors $\big((B_n, M_n)\big)_{n \in \mathbb{N}}$ is a Galton–Watson multitype branching process. Let Φ_n be the probability generating function associated with (B_n, M_n) for all $n \in \mathbb{N}$. In that case, equation (4.1) is simply the translation of the composition formula for branching processes, $\Phi_{n+1} = \Phi_1(\Phi_n)$ [12].

4.3.2. Example with a complex structure. In this section, we introduce an example that is a good illustration of the benefit of the symbolic approach. We want to compute the distribution associated with the number of 'Y-structures' (see Figure 8). The plant development model is given by Figure 9. The alphabet is $V = \{m, m', b, d\}$, where d represents a dead bud and the stochastic F0L-system can be easily deduced from Figure 9. We have $\pi_{b,d} = p$, $\pi_{b,mbbm'} = 1 - p$, $\pi_{m,m} = 1$, $\pi_{m',m'} = 1$ and $\pi_{d,d} = 1$. All the other components of π are equal to zero. Counting the number of 'Y-structures' is equivalent to counting the number of patterns $m'm$. Thus, the pattern of interest is $m'm$. As for the structural property of $DW_V^{\pi,n}[s]$, we have to break down the structure of an ORT in a way that highlights how Y-structures appear in the topology and how they
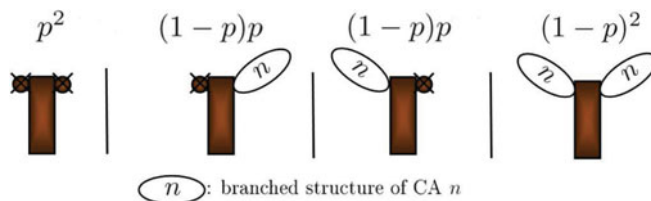
Figure 10. (Colour online) Decomposition of branched structures of CA $n + 1$.

are connected to substructures (see Figure 9). From Figure 9, we deduce that a word $w \in DW_V^{\pi,n+1}[b]$ can be:

- either a dead bud d with probability p,
- or a word of the form mvm' with $v \in DW_V^{\pi,n}[bb]$ with probability $1 - p$, in which case v represents a branched structure.

Let $\psi^n[b]$ be the GF associated with $m'm$ in $DW_V^{\pi,n}[b]$:

$$\psi^n[b](z) = \sum_{w \in DW_V^{\pi,n}[b]} (\Pi^n)_{b,w} \, z^{c(w,m'm)} = \sum_{k \in \mathbb{N}} P^{n,b}(k) z^k,$$

where $P^{n,b}(k)$ is the probability of getting k Y-structures in a plant generated by the organogenesis model described in Figure 9 after n GCs. Then, we have

$$\psi^{n+1}[b](z) = pz^{c(d,m'm)} + \sum_{v \in DW_V^{\pi,n}[bb]} (1-p)(\Pi^n)_{bb,v} z^{c(mvm',m'm)}$$

$$= p + (1-p)\psi^n[bb](z). \tag{4.2}$$

Now we need to find a recurrence relation for $\psi^n[bb](z)$. It can be deduced straightforwardly by the decomposition of Figure 10. We deduce that a word $w \in DW_V^{\pi,n+1}[bb]$ (*i.e.,* a branched structure) can be:

- two dead buds dd with probability p^2,
- one branched structure mvm' with $v \in DW_V^{\pi,n}[bb]$ and one dead bud d (either vd or dv) with probability $p(1-p)$ for each case,
- two branched structures mv_1m' and mv_2m' with $(v_1, v_2) \in DW_V^{\pi,n}[bb] \times DW_V^{\pi,n}[bb]$ with probability $(1-p)^2$.

Therefore,

$$\psi^{n+1}[bb](z) = p^2 z^{c(dd,m'm)}$$
$$+ \sum_{v \in DW_V^{\pi,n}[bb]} (1-p)p(\Pi^n)_{bb,v} \, z^{c(mvm'd,m'm)}$$
$$+ \sum_{v \in DW_V^{\pi,n}[bb]} (1-p)p(\Pi^n)_{bb,v} \, z^{c(dmvm',m'm)}$$
$$+ \sum_{(v_1,v_2) \in (DW_V^{\pi,n}[bb])^2} (1-p)^2(\Pi^n)_{bb,v}^2 \, z^{c(mvm'mvm',m'm)}.$$

Since

$$c(mvm'd, m'm) = c(dmvm', m'm) = c(v, m'm)$$
$$\text{and} \quad c(mvm'mvm', m'm) = 2c(v, m'm) + 1,$$

we obtain

$$\psi^{n+1}[bb](z) = p^2 + 2(1-p)p\psi^n[bb](z) + (1-p)^2 z(\psi^n[bb](z))^2. \tag{4.3}$$

By identifying the coefficients of the power series in equations (4.2) and (4.3), we get the distribution of Y-structures. The same method would work, for example, for computing the distributions associated with the number of apexes (*i.e.*, terminal nodes in a tree).

5. Application to parameter identification

In plant stochastic organogenesis models, the estimation of automata parameters may be complex, and necessitates important sampling work. For stochastic organogenesis models driven by a multitype branching process, classical methods based on branching processes have already been proposed (see [10] and [15]). These methods rely on the calibration of the parameters so that the theoretical mean and variance associated with the number of phytomers (computed from the model) are the closest to the experimental ones. However, phytomers are not always easily identifiable in a plant (no visible marks separating them after some time). Therefore, the idea is to use a botanical structure that is more easy to identify and to count (*e.g.*, apexes or Y-structures: see Section 4.3.2). By comparing the theoretical distribution of this structure to the experimental one, we are able to find the best set of parameters that will give the model the behaviour closest to real plants. However, this theoretical distribution can rarely be determined with classical branching process methods. In that case, the symbolic method of Section 4 provides a good alternative.

Let $\mathcal{S}$ be a stochastic organogenesis model driven by a multitype branching process and let $L = \langle W_V, \pi \rangle$ be the associated stochastic F0L-system. Suppose we are interested in computing the theoretical distribution of a particular structure in random plant architectures generated by $\mathcal{S}$. This particular structure is coded on the alphabet V by the word u. Let $\psi^n[s]$ be the GF associated with u in $SDW_V^{\pi,n}[s]$, where $s \in W_V$ is the word coding for a seed. As mentioned in Section 4.2, the theoretical distribution is given by the coefficient of $\psi^n[s]$ seen as a power series:

$$\text{for all } z \in [0,1], \quad \psi^n[s](z) = \sum_{w \in DW_V^{\pi,n}[s]} (\Pi^n)_{s,w} z^{c(w,u)} = \sum_{k \in \mathbb{N}} p^{n,s}(k) z^k.$$

Since $\text{card}(DW_V^{\pi,n}[s]) < \infty$ (the stochastic automata can only generate a finite number of structures), then, for all $n \in \mathbb{N}$, $\max\{c(w,u) | w \in DW_V^{\pi,n}[s]\}$ exists and is finite. In that case, we set:

$$\text{for all } n \in \mathbb{N}, \quad L^n = \max\{c(w,u) | w \in DW_V^{\pi,n}[s]\}.$$

Thus, for all $l > L^n$, $p^{n,s}(l) = 0$. Let ϕ^n be a vector in $[0,1]^{L^n+1}$ such that:

$$\text{for all } n \in \mathbb{N}, \quad \phi^n = (p^{n,s}(0), p^{n,s}(1), \ldots, p^{n,s}(L^n)).$$

Generally, the symbolic method leads to a recursive equation between $\psi^n[s]$ and $\psi^k[r]$ with $k \in K$ and $r \in R$, where K and R, respectively, are finite subsets of $\{0, \ldots, n-1\}$ and W_V. For the sake of clarity, we will suppose that the method gives us a recursive equation between $\psi^n[s]$ and $\psi^{n-1}[s]$ (the extension to the general case is straightforward). Therefore, by identifying the coefficients of the power series, we get a set of recurrence relations between the components of ϕ^n and ϕ^{n-1} which enables the building of ϕ^n from ϕ^{n-1}. The stochastic automata depend on a set P of parameters which have a botanical meaning (*e.g.*, the survival probability). As a consequence, the set of recurrence relations between the components of ϕ^n and ϕ^{n-1} also depends on P, and thus $\phi^n = \phi^n(P)$.

Suppose we have a plot of plants of CA $N \in \mathbb{N} \setminus \{0\}$. For each of these plants, we measure the number of patterns u. In so doing, we get the experimental distribution of the particular structure, $\{p^{\exp}(l)\}_{l \in \mathbb{N}}$. Suppose that there exists $l > L^N$ such that $p^{\exp}(l) \neq 0$. Then the organogenesis model is not well defined. In fact, for $l > L^N$, $p^{N,s}(l) = 0$. In that case, ϕ^N cannot get as close to the experimental distribution as we would like, and the organogenesis model (*i.e.*, the automata) needs to be modified. Then the model $\mathcal{S}$ is said to be well defined if

$$\max\{l \in \mathbb{N} | P^{\exp}(l) \neq 0\} \leqslant L^N.$$

Let $\phi^{\exp}$ be a vector in $[0,1]^{L^N+1}$ such that

$$\phi^{\exp} = (p^{\exp}(0), p^{\exp}(1), \ldots, p^{\exp}(L^N)).$$

The set of parameters P is estimated by the least-squares estimator $\hat{P}$,

$$\hat{P} = \underset{P \in [0,1]^{\mathrm{card}(P)}}{\mathrm{argmin}} \, \|\phi^{\exp} - \phi^N(P)\|_2^2,$$

where, for all $\alpha = (\alpha_0, \ldots, \alpha_{L^N}) \in [0,1]^{L^N+1}$,

$$\|\alpha\|_2 = \left(\sum_{i=0}^{L^N} (\alpha_i)^2\right)^{1/2}.$$

Several optimization algorithms can be used to find $\hat{P}$, such as the Levenberg–Marquardt algorithm (see [29]).

6. Conclusion

In this article, a symbolic method has been proposed to analyse the formation of complex structures in plants whose organogenesis is driven by a multitype branching process. For this purpose, a new combinatorial framework has been introduced. Plant structure is represented by an ordered rooted tree and can thus be coded by a Dyck word. The development rules of the organogenesis model are given by a stochastic 0L-system. In so doing, the development of plant structure is coded by a Markovian sequence of Dyck words. Therefore, studying plant structure and its development is completely equivalent to studying the Markovian sequence from a combinatorial perspective. A symbolic method was then established and this enables the computation of the distribution associated with the number of complex structures in plant topology.

This result could have interesting applications. For instance, in this article, we have shown how such a method can be used for the parameter identification of a stochastic organogenesis model, by comparing the theoretical distribution of a structure of a given type with experimental data.

For this purpose, we may take advantage of some interesting techniques that have recently been developed for automatic reconstruction of the morphology of individual trees from terrestrial laser scanner data [31, 37]. Such techniques may prove very useful in data acquisition, in order to evaluate the experimental distributions of the plant structural patterns of interest, from which the parameters from the organogenesis model can be estimated with our proposed methodology.

The type of stochastic organogenesis models considered in this paper assumes constant probabilities for the branching process underlying plant development. This is a reasonable hypothesis for plants with short life cycles, such as coffee or cotton plants, as well as plantation trees such as pine, poplar or eucalyptus trees [7]. However, the complex interactions between plant structural development and ecophysiological processes generate morphogenetic gradients [2, 26] that, in terms of modelling, should translate into variable probability laws evolving with plant aging or axis development [10]. The adaptation of the method to this case is a challenging subject.

Finally, the problem of pattern occurrences in a random text has been widely studied in various application fields, specifically molecular biology. Gene recognition and exceptional occurrence frequencies in DNA sequences are fundamental questions (see [30] or [33]). Since the proposed method applies to any multitype Galton–Watson branching process, it would be interesting to explore how it can be used in other research fields and application domains.

Acknowledgement

References

[1] Athreya, K. and Ney, P. (2004) *Branching Processes*, Dover.

[2] Barthélémy, D. and Caraglio, Y. (2007) Plant architecture: A dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. *Ann. of Botany* **99** 375–407.

[3] Bell, A. (1991) *Plant Form: An Illustrated Guide to Flowering Plant Morphology*, Oxford University Press.

[4] Capasso, V. and Morale, D. (2009) Stochastic modelling of tumour-induced angiogenesis. *J. Math. Biol.* **58** 219–233.

[5] Costes, E., Smith, C., Renton, M., Guédon, Y., Prusinkiewicz, P. and Godin, C. (2008) MAppleT: Simulation of apple tree development using mixed stochastic and biomechanical models. *Funct. Plant Biol.* **35** 936–950.

[6] Cournède, P.-H., Kang, M. Z., Mathieu, A., Barczi, J.-F., Yan, H. P., Hu, B. G. and de Reffye, P. (2006) Structural factorization of plants to compute their functional and architectural growth. *Simulation* **82** 427–438.

[7] Diao, J., de Reffye, P., Lei, X., Guo, H. and Letort, V. (2012) Simulation of the topological development of young eucalyptus using a stochastic model and sampling measurement strategy. *Computers and Electronics in Agriculture* **80** 105–114.

[8] Eichhorst, P. and Savitch, W. J. (1980) Growth functions of stochastic Lindenmayer systems. *Inform. Control* **45** 217–228.

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

[10] Guédon, Y., Barthélémy, D., Caraglio, Y. and Costes, E. (2001) Pattern analysis in branching and axillary flowering sequences. *J. Theoret. Biol.* **212** 481–520.

[11] Hallé, F., Oldeman, R. A. A. and Tomlinson, P. B. (1978) *Tropical Trees and Forests: An Architectural Analysis*, Springer.

[12] Harris, T. E. (1963) *The Theory of Branching Processes*, Springer.

[13] Hemmerling, R., Kniemeyer, O., Lanwert, D., Buck-Sorlin, G. and Kurth, W. (2008) The rule based language XL and the modeling environment GroIMP illustrated with simulated tree competition. *Funct. Plant Biol.* **35** 739–750.

[14] Henkel, A., Müller, J. and Pötzsche, C. (2012) Modeling the spread of phytophthora. *J. Math. Biol.* **65** 1359–1385.

[15] Kang, M. Z., Cournède, P.-H., de Reffye, P., Auclair, D. and Hu, B. G. (2008) Analytical study of a stochastic plant growth model: Application to the GreenLab model. *Math. Comput. Simul.* **78** 57–75.

[16] Karwowski R. and Prusinkiewicz, P. (2003) Design and implementation of the L+C modeling language. *Electron. Notes Theoret. Comput. Sci.* **86** 1–19.

[17] Kniemeyer, O., Buck-Sorlin, G. and Kurth, W. (2003) Representation of genotype and phenotype in a coherent framework based on extended L-systems. In *Advances in Artificial Life: Proceedings of the 7th European Conference on Artificial Life, ECAL* (W. Banzhaf, T. Christaller, D. Dittrich, J. T. Kim, and J. Ziegler, eds), Vol. 2801 of *Lecture Notes in Artificial Intelligence*, Springer, pp. 625–634.

[18] Kniemeyer, O., Buck-Sorlin, G. and Kurth, W. (2007) GroIMP as a platform for functional–structural modelling for plants. In *Functional–Structural Plant Modelling in Crop Production* (J. Vos *et al.*, eds), Springer, pp. 43–52.

[19] Knuth, D. E. (1997) *The Art of Computer Programming*, third edition, Addison-Wesley.

[20] Kurth, W. (1994) *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*, Berichte des Forschungszentrums Waldökosysteme der Universität Göttingen.

[21] Kurth, W. (1994) Morphological models of plant growth: Possibilities and ecological relevance. *Ecological Modelling* **75/76** 299–308.

[22] Lindenmayer, A. (1968) Mathematical models for cellular interactions in development I: Filaments with one-sided inputs. *J. Theoret. Biol.* **18** 280–289.

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

[24] Loi, C., Cournède, P.-H. and Françon, J. (2010) A symbolic method to compute the probability distribution of the number of pattern occurrences in random texts generated by stochastic 0l-systems. In *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms: AofA'10*, DMTCS Proceedings Vol. AM, pp. 473–488.

[25] Lopez, G., Favreau, R. R., Smith, C., Costes, E., Prusinkiewicz, P. and DeJong, T. M. (2008) Integrating simulation of architectural development and source-sink behaviour of peach trees by incorporating Markov chains and physiological organ function submodels into L-PEACH. *Funct. Plant Biol.* **35** 761–771.

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

[27] Mech, R. and Prusinkiewicz, P. (1996) Visual models of plants interacting with their environment. In *ACM SIGGRAPH'96 Proceedings*, pp. 397–410.

[28] Mode, C. J. (1971) *Multitype Branching Processes: Theory and Applications*, American Elsevier.

[29] Moré, J. J. (2006) The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis: Dundee 1977*, Vol. 630 of *Lecture Notes in Mathematics*, Springer, pp. 105–116.

[30] Pevzner, P. A., Borodovski, M. and Mironov, A. (1989) Linguistics of nucleotide sequences I: The significance of deviations from mean statistical characteristics and prediction of the frequencies of occurrence of words. *J. Biomol. Struct. Dyn.* **6** 1013–1026.

[31] Pfeifer, N., Gorte, B. and Winterhalder, D. (2004) Automatic reconstruction of single trees from terrestrial laser scanner data. In *Proceedings of 20th ISPRS Congress*.

[32] Prusinkiewicz, P. and Lindenmayer, A. (1990) *The Algorithmic Beauty of Plants*, Springer.

[33] Régnier, M. and Szpankowski, W. (1998) On pattern frequency occurrences in a Markovian sequence. *Algorithmica* **22** 631–649.

[34] Riordan, J. (2002) *An Introduction to Combinatorial Analysis*, Dover.

[35] Rozenberg, G. and Salomaa, A. (1980) *The Mathematical Theory of L-Systems*, Academic Press.

[36] Smith, A. R. (1984) Plants, fractals and formal languages. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, pp. 1–10.

[37] Xu, H., Gossett, N. and Chen, B. (2007) Knowledge and heuristic-based modeling of laser-scanned trees. *ACM Trans. Graphics* **26** #19.