Modeling Dependences among Plant Architecture Components using Multitype Branching Processes

P. Fernique J.-B. Durand Y. Guédon

CIRAD, INRIA(Virtual Plants, Montpellier; MISTIS, Grenoble), INRA, UM2

November 20, 2012

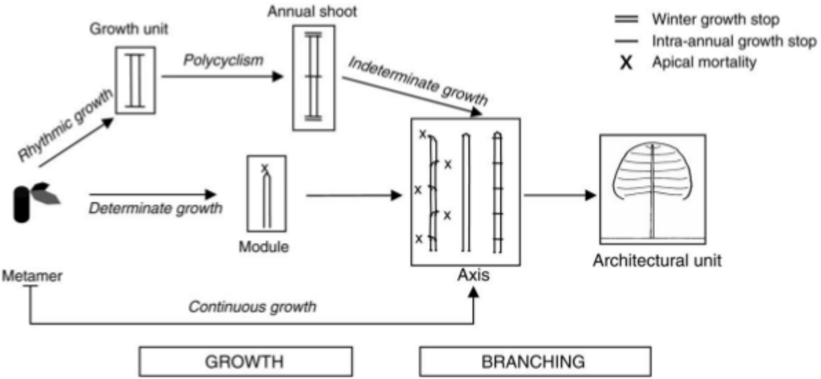
Quantitative biological analysis

Species selection, yield studies ... Considering important features in trees such as fruits where topology is important, one want statistical models taking into account various sources of structural fluctuations.

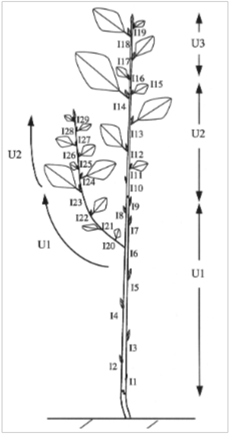
Testing biological hypothesis

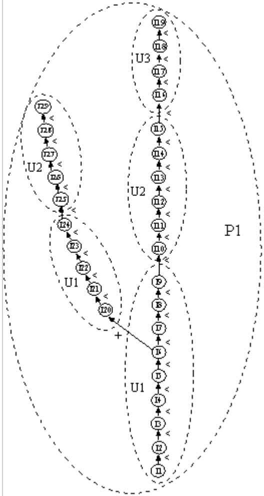
When using mechanistic models for important fluxes (IAA) in plants using L-systems one want to improve the simulation of the topology using statistical models of plant architecture inferred from training sets.

Model genesis
Plant growth [Barthélemy07]



Model genesis
Statistical individual [Godin98]

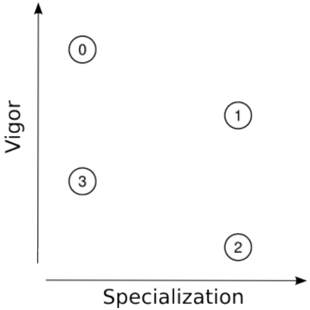




Each node having different properties: length, growth speed, number of flowers...

Coloration of the graph using states reflecting homogeneous characteristics of the GU

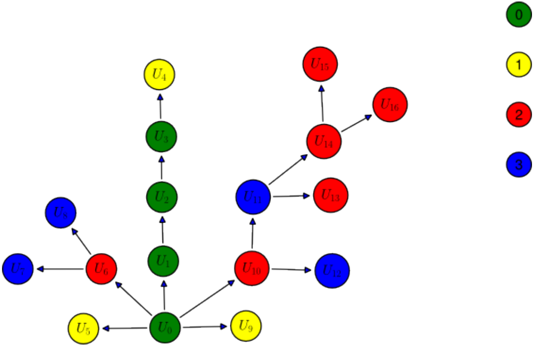
- Expert classification :
  - (0) Long and vegetative,
  - (1) Long and floral,
  - (2) Short and vegetative,
  - (3) Short and floral.



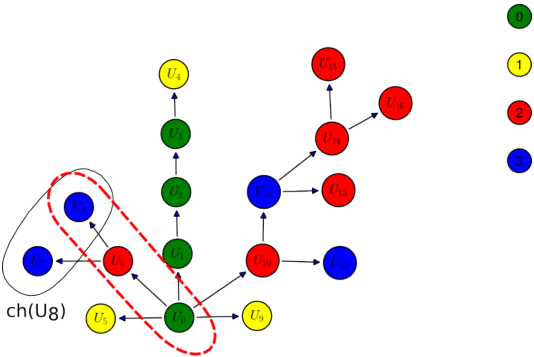
- Hidden Markov Models

Tree process distribution:

$$P[\mathbf{S} = \mathbf{s}]$$



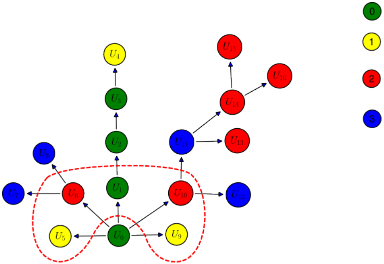
Markov hypothesis !



With limitation to local dependencies : Daughters shoots are independent from their non-descendants given their mother : $U_8 \perp\!\!\!\perp U_0 | U_6, U_1$ Which gives the following factorization :

$$P[\mathbf{S} = \mathbf{s}] = P[S_o = s_o] \prod_{u \in \mathcal{U}} P\left[\mathbf{S}_{ch(u)} = \mathbf{s}_{ch(u)} \middle| S_u = s_u\right]$$

PBMT are assuming that order is irrelevant : for a GU, only the number of children in each states is important.



For state O, one have the following histogram :

$$\left.\begin{array}{l}[1,2,2,0]:1\\ [1,0,0,0]:2\\ [0,1,0,0]:1\end{array}\right\}\mathbf{N}_u|S_u=0$$

PBMT are assuming that order is irrelevant:

$$P[\mathbf{S} = \mathbf{s}] \propto P[S_o = s_o] \prod_{u \in \mathcal{U}} P[\mathbf{N}_u = \mathbf{n}_u | S_u = s_u]$$

For many reasons :

- Order not always present in data,
- Order can be really hard to determine (less sampling effort),
- Is order really relevant ? (From non-order to partial order)

And only needs the specification of :

- 1 Initial distribution,
- K generation distributions

Where each generation distribution is a multivariate discrete distribution of K dimensions :

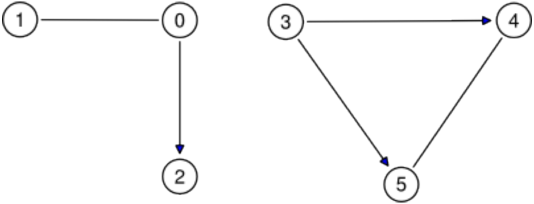
$$P[\mathbf{N}_u = \mathbf{n}_u | S_u = s_u]$$

Modeling dependences between children :

No dependences		All dependences
$\prod_{i=0}^{K-1} P\left[N_{u,i} = n_{u,i} \mid S_u = s_u\right]$	$\leftrightarrow$	$P\left[\mathbf{N}_u = \mathbf{n}_u \mid S_u = s_u\right]$
Covariances $= 0$		Covariances $< 0, > 0$

Research of distribution factorization where each component will be modeled using : parametric discrete (multivariate) distributions (regressions) for parsimonious models.

Partially Directed Acyclic Graph



$$\mathbf{N}_{0,1} \sim m - \mathcal{P}(0.2, 0.4, 0.1)$$

$$N_2 | N_1 = n_1 \sim \mathcal{B}(5, f(n_1))$$

$$N_3 \sim \mathcal{NB}(3.1, 0.3)$$

$$\mathbf{N}_{4,5} | N_3 = n_3 \sim \mathcal{M}(\mathcal{NB}(2., f(n_3)), g(n_3))$$

Number of vertices	2	3	4	5
Number of UGs	2	8	64	1,024
Number of DAGs	3	25	543	29,281

Structure learning :

- Restricted Family of graphs for which solutions can be easily found : Tree Graphs [Chow and Liu, 1968]
- Stepwise research with restricted search space by choice of a heuristic [Agresti, 2002; Koller, 2009]