

Segmentation-based approaches for characterising plant architecture and assessing its plasticity at different scales

Jean-Baptiste Durand⁽¹⁾, Yves Caraglio⁽²⁾, Patrick Heuret⁽³⁾, Eric Nicolini⁽²⁾

(1) Laboratoire Jean Kuntzmann – INRIA *Virtual Plants* – Grenoble Universités
51 rue des Mathématiques, BP 53, 38041 Grenoble Cedex 9, France
jean-baptiste.durand@imag.fr

(2) CIRAD, (3) INRA, UMR Botanique et Bioinformatique de l'Architecture des Plantes
TA A-51/PS2, 34398 Montpellier Cedex 5, France
{caraglio|nicolini|heuret}@cirad.fr

Keywords: hidden Markov tree model, categories of entities within tree structures, *Symphonia globulifera*, *Fagus sylvatica*

Introduction

Plants are modular organisms that develop by the repetition of elementary botanical entities or constructional units through the three main and fundamental morphogenetic processes of growth, branching and reiteration (Barthélémy and Caraglio, 2007). Repetition of these entities induces gradual or abrupt changes in their characteristics. These characteristics are quantified through several variables, called the entity attributes. On the one hand, differences between entities reflect different stages of differentiation in the meristems (Nicolini and Chanson, 1999), which are ordered in time and correspond to the notion of physiological age of meristems (Barthélémy and Caraglio, 2007). The changes of one attribute through the plant structure are referred to as morphogenetic gradient. On the other hand, part of the entity differences can be imputed to environmental factors.

Based on this botanical model of plant functioning, our approach jointly relies on categories of entities with similar characteristics, a description of those within-category characteristics and the category topological organisation within the plant. This constitutes a useful summary of the plant architecture, which is the basis for 1) highlighting hidden regularities in plant structures, for a better understanding of the gradients and organisation rules; 2) proposing and validating ecophysiological hypotheses, and new sampling protocols; 3) adopting a powerful paradigm for modelling plant development.

In this work, the entity categories are identified using a hidden Markov tree (HMT) model, in which local dependencies only (*i.e.* interactions between connected entities) are accounted for. This is not sufficient to accurately describe the topological distribution of the entities within the plant, and the changes of the plant topology. This is why various complementary methods and models (among which edit distance algorithms and sequence analysis using Markovian models), performed at different scales, are used in our approach to provide a more detailed description of the architecture, and to assess how various controlled factors affect architecture plasticity.

Segmentation of plant entities using hidden Markov tree models

Measurements collected on plants are coded as Multiscale Tree Graphs (MTGs), considered at a single given scale. The root vertex corresponds to the first entity at the basis of the tree trunk; thus the tree is oriented from the topological root vertex to the terminal (or “leaf”) vertices. The HMT model is used to perform entity segmentation and to represent the state transitions within tree-structures (see Durand *et al.*, 2005). HMT models are statistical models that assign a hidden state to each observed vertex. As a first step, the model parameters are estimated. Model parameters provide information about the state dynamics through the tree structure. The second step corresponds to a segmentation using the observed characteristics and the parameter values; this results into assigning a state value to each entity, taking into account connections between entities. This state is expected to reflect the stages of differentiation of the meristems, and can be

used as a summary of the entity attributes if the states are well separated. Those algorithms are implemented in the OpenAlea software (Pradal *et al.*, 2007).

The HMT model with conditionally independent children states, given the parent state, captures some local (parent-child) dependencies between tree entities. However, the HMT model is unable to account for direct dependencies between several ancestor vertices, or for changes, concerning state definition or dynamics, caused by individual variability (or other variability sources). On the basis of experiences with hidden Markov models for sequences, we expect the segmentation to be robust to these model limitations. This is related to certain robustness when estimating the distributions of the observed characteristics given the states, even if the dynamics of the states are poorly modelled. Such robustness can be interpreted as well-differentiated stages, in terms of entity characteristics.

Analysing the architecture of *Symphonia* and beech trees

This section presents a methodology based on entity segmentation for analysing the architecture of young *Symphonia globulifera* L. f. (Clusiaceae) and branching systems of beech trees (*Fagus sylvatica* L, Fagaceae). Firstly, categories of tree entities were obtained using an HMT model, which revealed the attribute distribution within each state and the probabilistic succession of the states within each individual. The states can be interpreted as stages of differentiation of the meristems.

Then the state frequencies were computed at the individual scale. Thus each tree was characterised by a vector of frequencies, which was used to perform a clustering of the individuals. The trees within one tree-cluster are expected to have comparable degrees of architectural development. The variability of the state frequencies according to controlled factors was also assessed.

At the scale of branching systems, the identification of borne complexes of similar importance was achieved by computing the edit distance between trees proposed by Ferraro and Godin (2000). This distance was computed using a local distance based on state labelling only, as opposed to the attributes. This corresponds to a denoising step where the attribute variability (for example due to variations in environmental conditions) is eliminated. Segmentation also avoids undesired interactions between local and global metrics. The question of synchronous growth of the trunk entities with those of the branching systems, or synchronous growth of several branching systems, was handled by extracting alignments provided by the edit distance computation.

Symphonia

30 young *Symphonia globulifera* individuals are considered. They were grown for 2 years in a greenhouse nursery in semi-controlled conditions under 3 light treatments of 5%, 10% and 20% of incident light. The aim of this study is to describe how *S. globulifera* individuals develop, and how their architecture is related to light environment (see Heuret *et al.*, in this workshop).

The HMT analysis was performed at the growth unit (GU) scale. The considered attributes are GU length, number of pairs of cataphylls and number of pairs of leaves. The model identification stage led to a seven-state model, where the states can be interpreted as levels of vigour, or basal position in borne branched systems.

The GU segmentation highlights some heterogeneity concerning the state frequencies, discriminating individuals with greater vigour and branching intensity than average (characterised by a high frequency of GUs in states of highest vigour) from those with less advanced development (low frequencies of those states). Moreover, the state frequencies were computed for each light treatment. As a result, low-vigour GUs have a markedly higher frequency for the 5% light treatment than for the other treatments, whereas GUs related to branching have lower frequencies.

Beech trees

A first set of 16 30-year-old beech trees is considered. This set is divided into 2 subsets of 8 individuals grown in different places. This is completed with a second set consisting of the last 25 years of growth of tree tops of 43 100-year-old individuals. The aim of this study is to highlight ordered stages in the differentiation

sequence of the meristems, and to assess the effect of population diversity (age and conditions of growth) on the number, the nature and the dynamics of the stages.

The HMT analysis was performed at the annual shoot (AS) scale. The considered attributes were AS length, number of cycles of extension and presence/absence of flowers. HMT models were identified separately on each dataset, and joined into a third model estimated using both datasets. This led to a twelve-state model.

On the basis of the AS segmentation, four clusters of individuals were visually identified in the first dataset: strong growth, moderate growth, slow growth and flowers. The state frequencies were compared between both growth places. This shows a clear separation between individuals with highest or lowest frequencies of the medium, short and flowering AS (corresponding to three given states). To validate the interpretation of the states as a quantification of physiological age, their stability (number and the nature) was assessed when extending the population diversity (age and conditions of growth). The distribution of the states along the trunk was also computed and compared with their global distribution, which highlights a higher frequency of long bicyclic AS than on the branches.

Moreover, the analysis performed on the second dataset emphasised complex patterns in the state succession leading to flowering, which could be characterised using variable-order Markov chains applied to selected paths within the trees. Synchronous nature of the AS on the trunk and branches was also emphasised by the states, which could be quantified through sequence alignments between the trunk and the borne axis.

We intend to carry other approaches to validate this methodology, by assessing how increasing the population diversity (for example diverse social status), or selecting different sets of attributes to build the HMT model, affects the state number or dynamics. The model ability to predict future flowering can also be investigated, by removing flowering AS from the data and identify some patterns characterising near apparition of flowers.

Importance of entity segmentation for assessing architecture variability at branching system or individual scales

Entity segmentation has a polyvalent status in the above studies: firstly, this provides a synthetic overview of the gradients, represented in the model through attribute distributions within each state, and the transition probabilities between states. Adding the state as a new synthetic variable in the data, allow complementary approaches for benefiting from a reduction of complexity (dimensionality), and some denoising. Secondly, the state-based approach reinforces the botanical notion of physiological age, since the state number remains very low, given the large complexity of measured trees, the number of individuals and of entities (hence the entity diversity). However, our model is not appropriate to model patterns, to relate the states with the topological growth, nor to resume the characteristics of whole branched systems. This is why HMT-based segmentation has to be combined with those complementary approaches, to precise and characterise states of architectural development.

Acknowledgments

The authors are thankful to Sabrina Coste for providing the plant material, and to Pascal Ferraro, Aïda Ouangraoua, Jean Bérard and Anne Perrut concerning the use of various tools for data analysis.

References

Barthélémy D. and Caraglio Y. 2007. Plant architecture: a dynamic, multilevel and comprehensive approach to plant form, structure and ontogeny. *Annals of Botany*, 99, pp. 375-407.

Durand J.-B., Guédon Y., Caraglio Y. and Costes E. 2005. Analysis of the Plant Architecture via Tree-structured Statistical Models: the Hidden Markov Tree Models. *New Phytologist*, 166(3), pp. 813-825.

Ferraro P. and Godin C. 2000. A distance measure between plant architectures. *Annals of Forest Sciences*, 57, pp. 445-461.

Nicolini E. and Chanson B. 1999. La pousse courte feuillée, un indicateur du degré de différenciation chez le Hêtre (*Fagus sylvatica* L.). *Canadian Journal of Botany*, 77(11), pp. 1539-1550.

Pradal C., Dufour-Kowalski S., Boudon F. and Dones N. 2007. The architecture of OpenAlea: A visual programming and component-based software for plant modeling. Submitted to FSPM07.