Proceedings of the 7th International Conference on Functional-Structural Plant Models, Saariselkä, Finland, 9 - 14 June 2013. Eds. Risto Sievänen, Eero Nikinmaa, Christophe Godin, Anna Lintunen & Pekka Nygren. http://www.metal.fi/fspm2013/proceedings. ISBN 978-951-651-408-9.

Deciphering mango tree asynchronisms using Markov tree and probabilistic graphical models

Anaëlle Dambreville[1,3], Pierre Fernique[2], Christophe Pradal[2], Pierre-Eric Lauri[3], Frédéric Normand[1], Yann Guédon[2] and Jean-Baptiste Durand[4,2]*

*[1]CIRAD, UPR HortSys, 97455 Saint-Pierre Cedex, Réunion Island, France, [2]CIRAD, UMR AGAP and INRIA, Virtual Plants, 34095 Montpellier, France, [3]INRA, UMR AGAP, 34098 Montpellier, France, [4]Grenoble University, Laboratoire Jean Kuntzmann, BP53, F-38041 Grenoble Cedex 9, France, *correspondence:* jean-baptiste.durand@imag.fr*

Highlights: Tree development is often characterised by complex dependencies between daughter growth units (GUs) deriving from a given mother GU, the so-called sister GUs. These dependencies directly affect the reproductive and vegetative phenological patterns that are at the origin of asynchronisms between adjacent GUs, eventually leading to within-canopy patchiness. These phenomena are rather common on tropical fruit-trees. We introduce new parsimonious statistical models to identify such dependencies. The proposed approach is illustrated on mango tree, a tropical species with a particularly complex timing of development. We focus especially on differences on fates and dates of burst between the daughter GUs issued from a same mother GU.

Keywords: growth asynchronism; growth unit; *Mangifera indica*; patchiness; spatio-temporal data analysis.

INTRODUCTION

As other tropical trees, mango tree is characterised by strong phenological asynchronisms between and within trees entailing patchiness (Chacko, 1986). Patchiness is characterized by clumps of either vegetative or reproductive development within the canopy: while some portions of the tree canopy develop vegetative GUs (i.e. portions of leafy axes developed during an uninterrupted period of growth), other portions may remain in rest or produce inflorescences at the same time. These asynchronisms often correspond to more or less large branching systems, e.g. scaffolds (Ramírez and Davenport, 2010). They entail various agronomical problems, such as the repeated use of pesticides to protect recurrent susceptible phenological stages from pests or a too extended period of fruit maturity, which may lead to difficulties to organize fruit harvesting.

If all terminal GUs produced both vegetative and reproductive daughter GUs in the same proportions and synchronously, i.e. at the same burst dates, all branching systems would grow synchronously and would have the same distribution of fates. Patchiness results from mutual exclusions, at the local scale of daughter GUs of a given mother GU, between some of their burst flushes (early, intermediate, or late) and / or some of their fates (vegetative or reproductive if it produces terminal or lateral inflorescences). These exclusions are observed, for example, when two kinds of daughter GUs cannot be produced by the same mother GU. Our final objective was to identify and characterise such exclusions and to open new perspectives to eventually connect them to patchiness at the canopy scale. Previous studies showed that the fate and burst date of a daughter GU are strongly affected by those of some ancestor GU (Dambreville *et al*., 2013). This approach, based on regression models, only made it possible to identify the effects of several factors (e.g. timing of development or fate of the mother GU, fruit load) on a single response variable, called GU feature (e.g. either the timing of development or the fate of a single daughter GU). This approach suffered from two main limitations: (i) multiple features of a GU cannot be predicted together in an obvious manner; (ii) a feature cannot be globally predicted for all daughter GUs if interactions exist between sister GUs, additionally to those with the mother GU.

The analysis of such interactions appears nevertheless essential to identify in which architectural and phenological contexts vegetative growth or flowering patches can occur. We present a new statistical methodology to reveal and describe these interactions using Markov tree and probabilistic graphical models (PGMs). Markov tree models allow the analysis of dependencies between a mother GU and its daughter GUs. They were introduced by Durand et al. (2005) to model plant architecture, assuming that sister GUs were independent given the mother GU. Here, the Markov tree model is generalised by introducing dependencies between sister GUs, in addition to their dependencies with their mother GU. The PMGs are used to unravel the dependencies between sister GUs, which express for example competition between these.

MATERIALS AND METHODS

The experimental orchard was located at the CIRAD (French Agricultural Research Centre for International Development) research station in Saint-Pierre, Réunion Island. Five trees of the Cogshall mango cultivar growing in the same plot were described at the GU scale (Dambreville *et al*., 2013). We defined the growing cycle as the period composed of the succession of a vegetative development period (appearance of new GUs) and a reproductive development period (flowering and possibly fruiting). Starting from the growing cycle of 2003, our experiment was carried out during the two following growing cycles in 2004 and 2005. We exhaustively described the development (fate, and in case of reproductive fate, number of terminal and lateral inflorescences, and burst date at the month scale) of all GUs appeared from sequential growth, i.e. located at the periphery of the canopy. Information on GUs burst in 2003 was limited to the fate, without distinction between terminal and lateral inflorescences for GUs that flowered. The 0.7% of GUs that burst after a two-year delay or more with respect to their mother GU were not considered in the study, and neither were their descendants.

The Markov tree models rely on three main assumptions: (i) the states of sister GUs (e.g. their fates) are independent from the states of their non-descendant GUs given the mother GU state (Markovian assumption); (ii) the joint distributions of the states of sister GUs are invariant under any permutation of the sisters (non-ordered sisters); (iii) these distributions do not depend on the position of the mother GU within the tree (homogeneity assumption). Let K denote the number of states and $\{0,\ldots,K-1\}$ the state space, let u be some non-terminal vertex and S_u its state variable. As a consequence of the three assumptions, the Markov tree model is entirely specified by the joint distributions of the numbers (N_0,...,N_{K-1}) of daughters with state 0,...,K-1 respectively, given the parent state $S_u=k$ (referred to as generation distributions).

Since the combinatorics induced by the variable number of sisters and the number of states is typically huge, the probability of occurrence cannot be reliably estimated by frequencies, and parsimonious parametric models must be used. In our approach, parsimony results from independence properties between the K count variables N_0,...,N_{K-1}, given $S_u=k$, summarised by a PGM; see Koller and Friedman (2009). The PGM vertices are the random variables $\{N_k\}$ and their connections correspond to conditional independence properties between the variables. Parametric models were obtained by combining discrete distributions and corresponding regression models (Fig. 1). Discrete distributions (chosen among Poisson, binomial, negative binomial and mixtures) were used to represent sources (vertices without parents) and the corresponding regression models were used to represent non-source vertices. To identify the PGM from data, an iterative algorithm was used to select edges, based on edit operations (deletion/insertion of edge, edge reversal). The PGM with maximum Bayesian information criterion (BIC) value was selected.

RESULTS AND DISCUSSION

To characterize dependencies (in particular, exclusions) between daughter GUs through their architectural and phenological context, it is necessary that the notion of GU state combines: (i) the period of growth of daughter GU, with respect to its mother GU burst date, i.e. immediate (I; i.e. during the same growing cycle) or one-year-delayed (D); (ii) the flush, i.e. early (E), Intermediate (I) or late (L); (iii) the fate, i.e. vegetative (V), reproductive with terminal flowering (T) or reproductive with lateral flowering (L). For GUs of 2003, the flush (U) and the position of flowering (F) were undefined. Thirteen states were defined for GUs as follows: U-V, IE-V, IL-V, DE-V, DI-V, DL-V, U-F, II-T, IL-T, DI-T, DL-T, II-L and DI-L.

Thus, 13 PGMs were identified, each one associated with one mother GU state. Since flushes were ordered (early, intermediate, late), and flowering mainly occurred for the last two flushes, states were partially ordered. As a consequence, the generation distributions simplified for states occurring late in the generation process, with respect to states occurring early.

We focus on the graph in Fig. 1 associated with the state II-L of the mother GU.

- No transition from the mother GU state II-L to the daughter GU states U-V, U-F, II-V, II-T, IL-V, IL-T, DE-V, nor II-L occurred. The states U-V and U-F (GUs produced in 2003) always preceded all the states. The other daughter GU states could theoretically follow II-L but this was not observed. This is translated in Fig.1 by isolated vertices with associated (quasi-) degenerate distributions for these eight states.
- The edges originating from source vertices DL-V and DL-T and pointing toward non-source vertex DI-V with associated negative regression parameters expressed mutual exclusion between DI-V on the one hand, and DL-V and DL-T on the other hand. The same mutual exclusion behaviour occurred between states DL-V and DI-L. This suggests that immediate intermediate GUs with lateral inflorescences (state II-L) could not have daughter GUs, the year after, successively at intermediate and then at late flushes.

Hence, mother GU in state II-L was a local context favourable to synchronism. This was consistent with the matrix of correlations between $\{N_k\}$ depicted in Fig. 2. However, despite a negative regression parameter associated with the edge pointing from vertex DI-T toward vertex DI-V, strict exclusion between both types of vertices did not strictly exist, since the correlation between the numbers of sister GUs in both states was non-negative (Fig. 2).

- The correlation matrix in Fig. 2 showed that the numbers of sister GUs in states DI-V and DI-L were quasi-uncorrelated, and that both are negatively correlated with the number of sister GUs in state DL-V. This was consistent with the selected PMG, which did not contain an edge between DI-V and DI-L.
- The numbers of sister GUs in states DL-T and DL-V were positively correlated, but it could be deduced from the PMG that this correlation was indirect, and was probably a consequence of their negative correlation with the number of sister GUs in state DI-V. This showed that daughter GUs with both vegetative (DL-V) and flowering (DL-T) fates may be produced synchronously, at flush L and with one-year delay, by mother GUs in the state II-L.

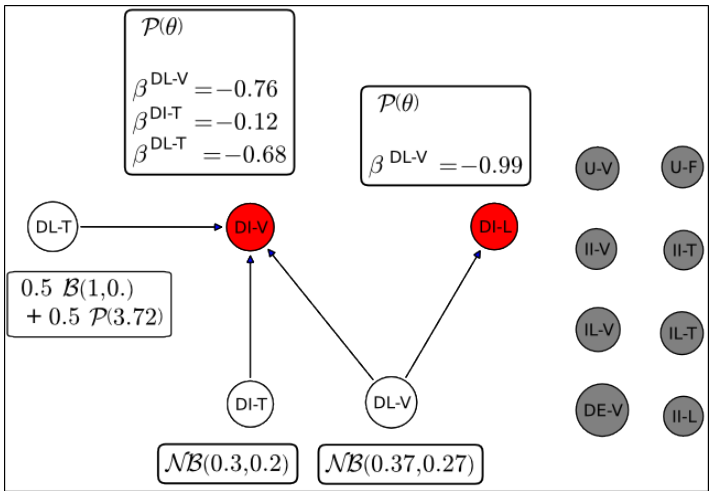

Fig. 1. PGM associated with the state II-L of the mother GU. Vertices of the PGM correspond to the random numbers of daughter GUs in each state. Grey vertices correspond to degenerate distributions. White vertices (source) and parameters correspond to univariate distributions ($\mathcal{P}$: Poisson, $\mathcal{B}$: binomial, $\mathcal{NB}$: negative binomial, and a mixture of $\mathcal{B}$ and $\mathcal{P}$ with weights 0.5 for vertex DL-T). Red vertices (non-source) correspond to univariate regressions. The parameter associated to the effect of variable i in the regression of the variable associated with a given vertex is denoted by β^i.

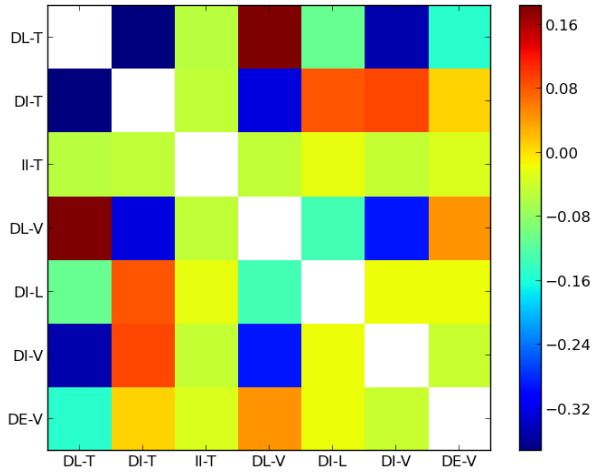

Fig. 2. Correlation matrix of the number N_k of daughter GUs in each state, given the mother GU state II-L (immediate intermediate flush with lateral flowering). Intensity of the correlation between N_j and N_k at location (j,k) is indicated using a colourmap. Blue tones correspond to negative correlations and red tones to positive ones. Some variables N_k with degenerate distributions are not represented.

These results show the ability of the Markov tree models to identify in which contexts a given mother GU can have or cannot have daughter GUs at different flushes or with different fates, which can be interpreted as the origin of asynchronism. This local point of view on asynchronism can be turned into a more integrated view by predicting, using our model, the total number of descendant GUs at each flush and each fate within a branching system, e.g. a scaffold, contributing to bring knowledge on the architectural determinants of patchiness. As a perspective, including explanatory variables into the generation distributions will allow to determine the effect of cultivar, growth conditions (climate) or horticultural practices (e.g. pruning or fruit thinning) on patchiness at different scales, and particularly at the whole-canopy level.

LITERATURE CITED

Chacko EK. 1986. Physiology of vegetative and reproductive growth in mango (*Mangifera indica* L.) trees. *Proceedings of the First Australian Mango Research Workshop*. CSIRO, Melbourne, 54–70.

Dambreville A, Lauri PÉ, Trottier C, Guédon Y, Normand F. 2013. Deciphering structural and temporal interplays during the architectural development of mango trees. *Journal of Experimental Botany doi 10.1093/jxb/ert105.*

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

Koller D, Friedman N. 2009. Probabilistic graphical models: principles and techniques. MIT press, 2009.

Ramírez F, Davenport TL. 2010. Mango (*Mangifera indica* L.) flowering physiology. *Scientia Horticulturae* **126**: 65-72.