Ilya Potapov, Marko Järvenpää, Markku Åkerblom, Pasi Raumonen and Mikko Kaasalainen

Data-based stochastic modeling of tree growth and structure formation


Highlights
- We propose a stochastic version of the tree growth model LIGNUM for producing tree structures consistent with detailed terrestrial laser scanning data, and we provide the proof-of-concept by using model-based simulations and real laser scanning data.
- Trees produced with the data-based model resemble the trees of the dataset, and are statistically similar but not copies of each other; the number of such synthetic trees is not limited.

Abstract
We introduce a general procedure to match a stochastic functional-structural tree model (here LIGNUM augmented with stochastic rules) with real tree structures depicted by quantitative structure models (QSMs) based on terrestrial laser scanning. The matching is done by iteratively finding the maximum correspondence between the measured tree structure and the stochastic choices of the algorithm. First, we analyze the match to synthetic data (generated by the model itself), where the target values of the parameters to be estimated are known in advance, and show that the algorithm converges properly. We then carry out the procedure on real data obtaining a realistic model. We thus conclude that the proposed stochastic structure model (SSM) approach is a viable solution for formulating realistic plant models based on data and accounting for the stochastic influences.

Keywords form diversity; morphological plasticity; terrestrial lidar; stochastic functional-structural plant model; quantitative structure models; data fitting

Addresses Tampere University of Technology, Department of Mathematics, P.O. Box 553, FI-33101 Tampere, Finland
E-mail ilya.potapov@tut.fi

Received 29 June 2015 Revised 14 October 2015 Accepted 14 October 2015

Available at http://dx.doi.org/10.14214/sf.1413
1 Introduction

Readily available algorithms for producing realistic and accurately data-calibrated synthetic tree forms would have numerous applications in botany and ecology. Thus far, studies of tree forms have focused either on data representations, with models used only to reconstruct a particular tree (stand) (Pfeifer et al. 2004; Rutzinger et al. 2011; Raumonen et al. 2013, 2015), or on theoretical models with estimated real conditions used for parameter values (De Reffye et al. 1988; Prusinkiewicz and Lindenmayer, 1990; Weber and Penn 1995; Perttunen et al. 1996). Our aim is to bridge this gap between data and theory. We use a stochastic tree growth model whose structural properties are iteratively fitted to the data to produce tree forms (not limited in number) that are statistically similar to the data tree or trees, but are not copies of each other.

To obtain computationally feasible growth models, we must make a twofold hypothesis in this study. First, we propose that a model can capture the variation of tree forms by a combination of simple deterministic processes and stochastic choices representable by low-dimensional probability distribution functions $p$. The model can thus reproduce tree shapes even if it is not exactly correct in the biological or probabilistic sense. Second, we assume that tree forms can be represented by sets of low-dimensional spaces with abstract coordinates $u$ in which the occurrence of various geometric features can be depicted, and that the model $p$ can be determined from field-data sets $U$ containing a number of observed $u$-points when both $p$ and $u$ are well chosen. Note that $p$ operates in the space of model parameters $q$, whereas $u$ represents the space of observables.

We thus need to create both forward and inverse mappings

$$p(q) \rightarrow U_{model}, \quad U_{data} \rightarrow p(q) \quad (1)$$

Once the former is defined, the latter is obtained by iteratively minimizing a distance measure $D_s(U_{model}, U_{data})$. The principle of introducing stochastic variation to the parameter values $q$ when the growth of a model is simulated can be applied to virtually any model: recursive (Honda 1971; Prusinkiewicz et al. 2001) or self-organizing (Ulam 1962; Bornhofen and Lattaud 2009); functional (Mäkelä and Hari 1986) or structural (Prusinkiewicz and Lindenmayer 1990; Fisher 1992), or any other.

2 Methods

The overall diagram of the method is depicted in Fig. 1. The main building blocks are: a quantitative structure model (QSM) containing all the geometric and hierarchical information about a tree, a stochastic functional-structural plant model (FSPM), a measure of distance between the structural data of the QSM and FSPM given in some $u$-spaces, and an optimization algorithm minimizing the distance (e.g., a genetic algorithm). The choices for the technical representations of any of the building blocks are not unique; these, in particular the choice of the distance measure and the $u$-spaces, will be discussed in greater detail elsewhere.

2.1 Data

All structural data can be obtained from a QSM derived from terrestrial laser scanning (TLS) measurements. We used the algorithm described in detail in Raumonen et al. (2013) to extract
a QSM from the TLS point cloud data. The field data used here were obtained at a plot of Scots pines in Ruotsinkylä, Finland in 2014 (mean height: 13.2 m, mean breast height diameter: 13.2 cm, stand density: 1932.8 trees·ha$^{-1}$, total volume: 196.9 m$^3$·ha$^{-1}$). Results reported here are based on a single QSM derived from the plot.

### 2.2 Model

As a FSPM to match the data tree structure, we use LIGNUM (Perttunen et al. 1996, 1998, 2001; Lu et al. 2011). Structurally the model consists of cylindrical segments. At each iteration of the model growth, the calculated overall carbon balance (photosynthetic production vs. respiration needs) and the light conditions determine how much biomass is allocated in the consecutive iteration.

The full radiation model implemented in LIGNUM (Perttunen et al. 1998) would be computationally expensive given the large number of iterative simulations necessary in our approach. Therefore, we have modified the radiation model using the shadow propagation method (Palubicki et al. 2009). We divide the space into a grid of voxels each with the associated shadow value $S(x,y,z)$. Each segment is ascribed to a particular voxel creating a pyramidal penumbra at the voxels underneath. Shadow propagation length $S_L$ and voxel edge size $V_S$ determine, how many voxel layers the shadow propagates down (Table 1; Palubicki et al. 2009). At the top voxel $S(x,y,z) = S_{\text{max}}$, whereas at the bottom $S(x,y,z) = 0$ (linear interpolation between).

All segments are sequentially processed, producing a 3D grid of accumulated shadow values $S(x,y,z)$. The radiation intercepted by a segment occupying the $(x,y,z)$ voxel is

$$I(x,y,z) = \max\left[0, R - S(x,y,z)\right]$$

(2)
and the photosynthetic light ratio is

\[ I_p(x, y, z) = I(x, y, z) / R, \]  

where \( R \) is the photosynthetically active radiation in the fully exposed conditions (Table 1). These determine the growth rate (Perttunen et al. 1996, 1998). Also, branches are not allowed to grow further than a radius \( R_{\text{env}} \) away from the tree stem, nor intersect other branches or the ground.

Branches with no foliage and no child branches are shed (Sievänen et al. 2008). Since real pine trees contain dead branches, which cannot be effectively distinguished from the live ones in the TLS set-up, we have introduced a delay factor: the youngest segment must be \( T_{\text{sh}} \) iterations old for the shedding to occur. Additionally, real trees might have branches shed only partially leaving the dead remnants due to accidental branch break, so we have introduced a Poisson random variable (parameter \( \mu_{\text{sh}} \)) determining the number of the tip segments (thinner and thus easier to break than the base) that will be shed while the rest will stay (Table 1).

### 2.3 Tree form description and model fitting

We determine several quantitative characteristics describing the tree shape, and plot their values as sets \( U \) in each chosen abstract plane of coordinates \( u \) as shown in Figs. 2 and 3. The variations of the density of the points in \( u \) are taken to be the QSM characteristics reflecting the stochastic nature of real tree growth.

The original LIGNUM parameters can be changed into and augmented with probability distributions (here simple Gaussian or Poisson ones; see Tables 1 and 2) to catch the stochasticity of growth. As the model tree grows, the effective LIGNUM parameter values \( q \) are repeated samples drawn from these distributions \( p \) defined by some parameters \( w \).

Using the point plots (Figs. 2 and 3), we compare the characteristic \( U_{\text{data}} \) and \( U_{\text{model}} \) of QSM and LIGNUM, respectively. The measure of structural distance \( D_s[U_{\text{data}}, U_{\text{model}}(w)] \) between the

### Table 1. The LIGNUM parameters used in this study. Parameters fixed during simulations are reported; the varied (-) and not used (NA) parameters are omitted. Other parameters are from Perttunen et al. (1998) for the synthetic simulation (Section 3.1) and Sievänen et al. (2008) for the real case study (Section 3.2).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Synthetic</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>full exposure radiation for a tree segment, relative</td>
<td>30.0</td>
<td>60.0</td>
</tr>
<tr>
<td>( S_{\text{max}} )</td>
<td>maximum shadow induced by a tree segment, relative</td>
<td>10.0</td>
<td>-</td>
</tr>
<tr>
<td>( S_L )</td>
<td>shadow propagation distance, [m]</td>
<td>0.55</td>
<td>-</td>
</tr>
<tr>
<td>( V_S )</td>
<td>linear voxel size, [m]</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>( R_{\text{env}} )</td>
<td>radius of the circular boundary around the tree stem that the branches are not allowed to cross, [m]</td>
<td>1.33</td>
<td>-</td>
</tr>
<tr>
<td>( L_R )</td>
<td>ratio between the radius and length of a tree segment, dimensionless</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( Q )</td>
<td>apical dominance parameter, varied between 0 (no dominance) and 1 (maximum dominance), dimensionless</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \beta_{\text{init}}, \beta_{\text{max}} )</td>
<td>initial and maximal inclination angles, respectively: angles between the first branch segment and the segment it emanates from, [degree]</td>
<td>- , 95</td>
<td>-</td>
</tr>
<tr>
<td>( \Delta \beta )</td>
<td>inclination angle increment, [degree]</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \Delta \zeta, \Delta \gamma )</td>
<td>intensity of the white noise added to the vertical (( \Delta \zeta )) and horizontal (( \Delta \gamma )) orientations of the segments, [degree]</td>
<td>- , 5</td>
<td>-</td>
</tr>
<tr>
<td>( T )</td>
<td>number of iterations/years of simulation</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( T_{\text{sh}} ) when shedding a branch, the lower age limit for the branch segments</td>
<td>NA</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( \mu_{\text{sh}} )</td>
<td>parameter to the Poisson variable determining number of the tip segments to shed</td>
<td>NA</td>
<td>-</td>
</tr>
</tbody>
</table>
Fig. 2. Sets of points describing the stochastic structure of a tree: a) the simulated LIGNUM tree, source for the data sets; b) the branching/inclination angles of all branches of the 2nd order, (1D set) and c) 1st order tapering, the local cross-section area radius of a branch against the distance from the base of the branch, gathered from all branches of order 1 (2D set). Order is the Gravelius order of a tree such that the lowest order (tree stem) equals to zero.
two is minimized by adjusting the parameters \( w \), as well as the deterministic LIGNUM parameters, in an iterative optimization procedure (here a genetic algorithm to enable global optimization). The \( D_s \) measures the difference between the local densities of the data and model points \( U \) in the chosen plot planes. Here its construction is based on Kaasalainen (2008) (cf. Sect. 5 and Eq. (32) therein), but this (or the \( u \)-spaces adopted here) is by no means a unique or optimal choice. For example, if we have a large number of \( u \)-points from a plot of clones or trees otherwise labeled similar, we can divide the \( u \)-space into cells, and minimize the model vs. data difference of the relative density of occupation numbers in each cell.

We call the set of resulting best-fit stochastic FSPM parameters a stochastic structure model (SSM). Note that this is not a fixed tree form: its computer realizations have each time different details but they share a statistically similar appearance.
3 Results

We report two cases, based on the synthetic (generated by stochastic LIGNUM) and real data, respectively. The former demonstrates the possibility of the algorithm to converge to the target tree form and the target parameter values that are known in advance. Additionally, it reveals practical aspects of the algorithm performance and assesses sensitivity of the parameters. Thus, the actual values of the parameters play no significant role in this case.

We opted for the 8-parameter estimation, which was found to be a compromise between the complexity and number of parameters to converge: as the number of parameters grows the possibility of various parametric solutions producing similar tree structures increases. The real data case has no such limitations.

3.1 Model-based synthetic data

As an initial consistency check and a test of our basic modeling principle, we fit the model to simulated QSM data generated by the same model (since the model is stochastic, we do not commit an “inverse crime”; i.e., the optimized SSM tree cannot look exactly like the original one). Here we select eight parameters (Table 2) to be estimated and fix others as in Perttunen et al. (1998). We set the parameter variation ranges for the optimization routine from 50% to 150% of the target value for every parameter, which helps us to assess the parameter estimation accuracy. The coordinates in which the fitting is depicted are obtained from the branches of the 1st Gravelius order (the stem is of order 0): we choose tapering and spatial curvature of the branches. The former couples the local radius of the branch and the distance from its base. The latter relates the two spatial angles (in horizontal and vertical planes) between the consecutive segments of a branch to the relative distance from its base. The sets $U$ are rendered as scatter plots by considering all branches of the given order. The results are shown in Fig. 3 and Table 2.

As Fig. 3 shows, the structure and shape characteristics of the data and model trees match well. Model characteristics such as tree height, trunk width, and crown branch distribution are close to those of the original simulated tree. Note that the model is not supposed to look exactly like the original tree, rather, to be consistent with it and share its main structural tendencies and stochasticity. Comparing the original and iterated parameter values, we can check whether the

### Table 2. Parameter estimation in the emulated “data” case. The tree shapes and fitted structural data sets are shown in Fig. 3. Parameters $L_R$ and $Q$ follow the normal distribution having two parameters: mean and standard deviation (std). Other LIGNUM parameters are from Perttunen et al. (1998) and $R = 30.0$, $S_{max} = 10.0$, $S_L = 0.55$ m, $V_S = 0.02$ m, $R_{env} = 1.33$ m, $\beta_{max} = 95$ degree, $\Delta \gamma = 5$ degree, and original shedding options (Sievänen et al. 2008). See Table 1 for the parameter definitions. The genetic algorithm has stopped after 22 full iterations/generations, each generation consisted of 40 parameter sets.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Data value</th>
<th>Model value (estimated)</th>
<th>Relative error, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_R$, mean</td>
<td>0.009</td>
<td>0.0094</td>
<td>4.65</td>
</tr>
<tr>
<td>$L_R$, std</td>
<td>0.001</td>
<td>0.0009</td>
<td>6.95</td>
</tr>
<tr>
<td>$Q$, mean</td>
<td>0.2</td>
<td>0.2058</td>
<td>2.88</td>
</tr>
<tr>
<td>$Q$, std</td>
<td>0.03</td>
<td>0.0213</td>
<td>28.88</td>
</tr>
<tr>
<td>$T$</td>
<td>15</td>
<td>16</td>
<td>6.67</td>
</tr>
<tr>
<td>$\Delta \bar{\beta}$</td>
<td>10.0</td>
<td>9.7665</td>
<td>2.33</td>
</tr>
<tr>
<td>$\Delta \zeta$</td>
<td>5.0</td>
<td>4.3708</td>
<td>12.58</td>
</tr>
<tr>
<td>$\beta_{init}$</td>
<td>35.0</td>
<td>40.4649</td>
<td>15.61</td>
</tr>
</tbody>
</table>
optimization indeed implies that the inverse relation \( D_s \approx 0 \Rightarrow p_{\text{model}} \approx p_{\text{data}} \) (uniqueness and stability of the inverse problem) holds.

Most of the fitted parameters converged closely to the corresponding target values as assessed by the relative error. However, the parameters \( \text{std } Q \) (standard deviation of the normally distributed parameter \( Q \)), \( \beta_{\text{init}} \), and \( \Delta \zeta \) have errors of estimation larger than 10\% (Table 2). This can occur because of several reasons. First, it can be poor performance of the optimization algorithm. Second, it can be due to the low sensitivity of the chosen characteristic \( u \) to the parameter variation so that the overall progress of the fitting does not strongly depend on the changes in these parameter values (i.e., the inverse problem is unstable). This, perhaps, occurs for \( \text{std } Q \), which affects the tree branches of order higher than 1 for the given value ranges. Finally, some parameters can contribute to the poorer convergence of the aforementioned parameters.

However, poorer parameter convergence does not necessarily mean a poorer result: what we want is a close match between \( U_{\text{model}} \) and \( U_{\text{data}} \), and it is possible that this can be achieved with nonunique LIGNUM parameter combinations. This is even more probable with a model with many parameters, and has also biological grounds: the processes leading to a particular appearance or structure of a complex organism are probably not unique. Simulations with 16 parameters to estimate are also in favor of this argument (data not shown). The real test is not so much the closeness of the original and iterated parameters, but the similarity of the data and model \( U \)-distributions shown in Fig. 3 and, ultimately, the visible similarity of the data and model trees. Further study is needed to develop criteria for the best parameter and \( u \)-plot plane choice of the model to be estimated.

### 3.2 Real data

Next, we perform a similar optimization procedure on real tree data rendered as a \( QSM \). The structure of the data tree is shown in Fig. 4a (leftmost). The LIGNUM parameters for modeling were chosen as in Sievänen et al. (2008), with 16 free parameters to be estimated. Additionally, the model included the stochastic shedding properties defined by the parameters \( T_{\text{sh}} \) and \( \mu_{\text{sh}} \) (Table 1). The target \( u \) plots were the tapering, the spatial curvature, and the branching angle, all for the 1st order branches. The results of this study are shown in Fig. 4.

First, we can see that there are certain drawbacks in the data, e.g., gaps between cylinders. Additionally, due to occlusion and distance effects the apical part of the crown is not fully captured by TLS; i.e., the crown appears to have fewer branches than in reality. Because of this, the coarseness of the model, and lack of the \( u \) plots representing the trunk information, the tree height was underestimated (error is about 7\%), whereas the crown branch density and the trunk base diameter were overestimated (error for the breast height diameter estimation is about 55\%). Nevertheless, the structure of the crown was decently captured, indicating that the choice of \( u \), the parameters to estimate, and the model are sufficient for the purpose. Note also five different sample “clone” trees all simulated with the same best-fit parameters. The trees represent similar overall structure with varying details of organization emulating the intra-species diversity (Fig. 4a).

The dead branches are also present in the model, and their length profile corresponds to that of the real tree. This is due to the ad hoc stochastic shedding rules included in LIGNUM (\( T_{\text{sh}} \) and \( \mu_{\text{sh}} \)). Although the measurements cannot provide for the foliage distribution in the tree, the model indicates that the foliage is spread at the upper part of the crown (Fig. 4b), which is plausible, given the natural conditions of this tree corresponding to the high-density tree stand.
4 Discussion

We have introduced the concept of Stochastic Structure Models (SSMs), based on the QSMs of real trees, and provided the proof-of-concept by simulations and real data. To make the SSM procedure capable of producing realistic general models of tree growth, the next steps will be to calibrate the models with a number of QSMs of trees of varying ages, include stochastic versions of fully synthetic (rather than biologically-based) flexible algorithms of structure formation such as those in Palubicki et al. (2009), experiment with various ways of introducing stochastic distributions in the synthetic or FSPM models, model various tree species, and include the effects of inter-tree competition required for the construction of the SSMs of entire forests. For example, the choices of $D_s$, the stochastic model parameters, and the coordinate planes $u$ are not unique and may depend on the species (and are in any case somewhat simplified in this initial study). Also, controlled experiments should be carried out to test the hypotheses and assumptions behind $D_s$, $u$, and $p$. For example, one could expect that cloned trees grown in similar circumstances should have smaller mutual $D_s$ than others, and the $D_s$ of clones should correlate with environmental variation (e.g., provenance studies).

We have introduced our concept to initiate an extensive program that requires a considerable amount of experimenting by a sizable collaboration network. For this purpose, our source code is downloadable from SSMLignumSF (2015). We emphasize that the technical choices in the code and the examples here are just one sample of a large number of options. To speed up the computations, parallel computing is highly recommended: this is easy to carry out, e.g., for the members of the trial populations used in the genetic algorithm. The optimization can take tens of hours of CPU time, so a cluster or cloud implementation reduces the real-time computational cost considerably.

The program code used in this work is available at: [http://github.com/inuritdino/SFennica-Lignum-SSM](http://github.com/inuritdino/SFennica-Lignum-SSM).
Acknowledgments

We are thankful to Sanna Kaasalainen for providing us with the TLS data and Risto Sievänen, Jari Perttunen, and Wojtek Palubicki for the valuable discussion on implementing the stochastic version of LIGNUM.

References


Prusinkiewicz P., Mündermann L., Karwowski R., Lane B. (2001). The use of positional informa-


Total of 22 references.