

# Bayes Forest: a data-intensive generator of morphological tree clones

Ilya Potapov<sup>1</sup>, Marko Järvenpää<sup>2</sup>, Markku Åkerblom<sup>1</sup>, Pasi Raumonen<sup>1</sup>, Mikko Kaasalainen<sup>1</sup>

<sup>1</sup>Mathematics Department, Tampere University of Technology, P.O. Box 553, 33101, Tampere,

Finland

<sup>2</sup>Helsinki Institute for Information Technology, Department of Computer Science, Aalto University,

Finland

Emails (in the author order, \* - corresponding author):

\* ilya.potapov@tut.fi, marko.j.jarvenpaa@aalto.fi, markku.akerblom@tut.fi, pasi.raumonen@tut.fi,  
mikko.kaasalainen@tut.fi

## Abstract.

**Background.** Detailed and realistic tree form generators have numerous applications in ecology and forestry. For example, varying morphology of trees contribute differently to formation of landscapes, natural habitats of species, and eco-physiological characteristics of the biosphere. Additionally, virtual clones might be used in studies of real (e.g. genetic) clones.

**Findings.** Here, we present an algorithm for generating morphological tree “clones” based on the detailed reconstruction of the laser scanning data, statistical measure of similarity, and a plant growth model with simple stochastic rules. The algorithm is designed to produce tree forms, i.e. morphological clones, similar as a whole (coarse-grain scale), but varying in minute details of organization (fine-grain scale). Although we opted for certain choices in our algorithm, individual parts may vary depending on the application, making it a general adaptable pipeline. Namely, we showed that specific multi-purpose procedural stochastic growth model can be algorithmically adjusted to produce the morphological clones replicated from the target experimentally measured tree. For this, we developed a statistical measure of similarity (structural distance) between any given pair of trees, which allows for the comprehensive comparing of the tree morphologies in question by means of

26 empirical distributions describing geometrical and topological features of a tree. Finally, we developed  
27 a programmable interface to manipulate data required by the algorithm.

28 **Conclusions.** Our algorithm can be used in variety of applications for exploration of the morphological  
29 potential of the growth models (both theoretical and experimental), arising in all sectors of plant  
30 science research.

32 **Keywords:** quantitative structure tree model; morphological clone; stochastic data driven model;  
33 terrestrial laser scanning; large scale data; empirical distributions; distribution tomography

## 35 Findings

### 37 I. Background

39 Models for plant architecture attract significant attention due to their ability to assist the empirical  
40 studies in ecology, plant biology, forestry, and agronomy [1]. The modeling activity is especially  
41 useful in research since it arises as fruitful collaboration between specialists in different fields of  
42 studies: computer scientists, mathematicians, and biologists [2].

44 Modeling plant architecture is approached from many directions. Some progress has been achieved in  
45 synthesis of realistic plant forms in the field of computer graphics [3-5]. These models, although based  
46 on heuristic rules of growth, produce realistic shape outcomes in a fast and efficient manner, which is  
47 usually dictated by the application of this approach, that is natural sceneries in computer visualization.  
48 Heuristic growth rules of the procedural models for graphics applications are not firmly based on  
49 biological principles, but nevertheless elucidate some algorithmic properties of the growth process (for  
50 example, recursive [6] vs. self-organizing [3, 7] character of architecture development).

51

1  
2 52 However, the most promising plant architectural models are so called functional-structural plant  
3  
4  
5 53 models (FSPM), also known as “virtual plants” [8-10], because this type of models allows for a  
6  
7 54 balanced description between morphological and functional/physiological properties of a plant. Thus,  
8  
9  
10 55 it is capable of connecting the external abiotic factors (e.g. radiation, temperature and soil) and the  
11  
12 56 most vital functions of a plant organism (such as photosynthesis, respiration, and water and salts  
13  
14 57 uptake) with its structural characteristics [1, 2].  
15

16  
17 58

18  
19 59 Nevertheless, biologically relevant architectural plant models rely on data in a form of empirically  
20  
21  
22 60 fitted functions and parameters that correspond to a particular species and/or certain site conditions  
23  
24 61 [11-14]. Thus, the change in these conditions requires re-calibration of the models, which is done in a  
25  
26  
27 62 manual fashion every time the model is simulated for the new conditions. Strong dependence on data,  
28  
29 63 where each simulation would be calibrated automatically by data, is limited by both computation time  
30  
31  
32 64 and lack of the fast measurement and processing systems allowing for a detailed 3D morphological  
33  
34 65 reconstruction of the real plant/tree.  
35

36 66

37  
38  
39 67 The most recent advances in laser scanning techniques allow for fast and non-destructive measurement  
40  
41 68 of trees with subsequent reconstruction of various characteristics depending on application (e.g. [15,  
42  
43  
44 69 16]). Most of such studies dedicated to reconstruction of 3D point clouds obtained from laser scanning  
45  
46 70 measurements deal with overall characteristics, such as height, width, and volume of stems/crowns,  
47  
48  
49 71 leaf index, biomass etc., resembling traditional destructive methods of measurement [15, 17].  
50  
51 72 However, the detailed precise geometrical and topological reconstruction with the preserved tree  
52  
53 73 architecture as is, is rarely sought after.  
54

55  
56 74  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 75 We use a fast, precise, automatic, and comprehensive reconstruction algorithm initially presented in  
2 76 [18] and further developed and tested in [19]. The algorithm reliably reconstructs a quantitative  
3  
4 77 structure model (QSM), which contains all geometrical and topological characteristics of the object  
5  
6  
7 78 tree. Input for the method is the 3D point cloud, sufficiently covering the tree, obtained from the  
8  
9  
10 79 terrestrial laser scanning measurements (TLS) and no additional allometric relations used for  
11  
12 80 estimation of the branch proportions (as in [20, 21]) are needed. Compared to other similar techniques  
13  
14 81 (e.g. [20-22]) this method requires few parameters and no user interaction and reconstructs the tree  
15  
16  
17 82 surface with subsequent cylinder (or any other geometrical primitive) approximation, which is usually  
18  
19 83 consistent with theoretical plant growth models. The reconstruction algorithm has been validated in  
20  
21  
22 84 several studies with several different tree species and different scanner instruments [19, 23-26]. There  
23  
24 85 are other published QSM reconstruction methods from TLS data that can produce similar quality  
25  
26 86 QSMs, at least [23].

27  
28  
29 87  
30  
31  
32 88 In this work, we utilize an inverse iterative procedure to optimize model's parameters as to match the  
33  
34 89 (empirical) distribution of structural features of the simulated stochastic tree models (FSPM, graphical  
35  
36 90 or other) to that of the tree reconstructed from the laser scanning data. Meanwhile, we formulate a  
37  
38  
39 91 measure of similarity of the tree structures grounded in tomographic analysis of the structural  
40  
41 92 distributions (e.g. Radon transform) [27, 28]. Finally, the optimal parameter set produces  
42  
43  
44 93 morphological "clone" trees with similar overall structure, but varying minute details of organization.

45  
46 94  
47  
48  
49 95 Recently, we have reported a proof-of-concept study where we used reconstruction of a pine tree and  
50  
51 96 the corresponding FSPM (named LIGNUM [13, 29]) to demonstrate the practical feasibility of the  
52  
53 97 approach [30]. Here, however, we develop a unifying interface (in the form of a programmable  
54  
55  
56 98 toolbox) for our procedure and use general-purpose fast procedural tree growth model from [3], since  
57  
58 99 such a simple procedural model is easier to adapt (it is simple, fast, and efficient) for technical  
59  
60  
61  
62  
63  
64  
65

100 experimentation with the whole algorithm. Similar algorithmic pipeline was reported in [5] for  
1  
2 101 procedural tree growth models in the context of graphics synthesis. However, in our approach we see  
3  
4  
5 102 the tree growth as a random process and, consequently, apply corresponding statistical methods for  
6  
7 103 measuring the similarity between trees. Moreover, in our algorithm the special concern is on  
8  
9  
10 104 biologically and physically relevant descriptions, hence, the careful choice of the reconstruction  
11  
12 105 algorithm; possibility to use FSPM to relate physiological parameters to the morphogenetic processes  
13  
14  
15 106 in trees; and no extra structures improving visual properties of trees but not supported by empirical  
16  
17 107 observation (e.g. leaves). Finally, any other choices of parameters and feature descriptions can be used  
18  
19 108 in our approach, further facilitated with a programmable interface.  
20  
21

22 109

## 24 110 **II. Algorithm overview**

26  
27 111

29 112 Our approach is based upon five distinct parts:

- 31  
32 113 1. *Quantitative Structure Model (QSM)* is a reconstruction of a tree model from 3D point clouds  
33  
34 114 obtained from terrestrial laser scanning measurements (TLS). Here we use specific algorithm for  
35  
36 115 such reconstruction reported in [18] and [19] but others could be used as well.
- 37  
38  
39 116 2. *Stochastic Structure Model (SSM)* is a tree growth model that is chosen depending on the  
40  
41 117 application. There are no limitations on the class of the model, except it must produce measurable  
42  
43  
44 118 3D branching structure.
- 45  
46 119 3. *Structural data set (U)* is a collection of structural features (empirical distributions) to be  
47  
48  
49 120 compared between QSM and SSM. Importantly,  $U$  data sets must be determined in the same way  
50  
51 121 both for QSM and SSM.
- 52  
53  
54 122 4. *Measure of structural dissimilarity, or structural distance  $D_S$* , is a measure of discrepancy between  
55  
56 123 any two data sets, in other words,  $D_S(U_1, U_2)$  results in a value quantifying how much different the  
57  
58 124 two data sets  $U_1$  and  $U_2$  are.  
59

60  
61  
62

125 5. *Optimization algorithm* is a numerical routine capable of finding a minimum of any given function  
1  
2 126 by varying its arguments (Newton algorithm, genetic algorithm, simulated annealing etc.)  
3  
4  
5 127  
6  
7 128 The connection between these components is outlined in Fig. 1 with explanation in the text below.  
8  
9  
10 129  
11  
12  
13 130 **Figure 1: The algorithm outline (see explanation in the text).**  
14  
15  
16 131  
17  
18 132 The algorithm outline (Fig. 1):  
19  
20  
21 133  
22  
23 134 Preparation stage A:  
24  
25 135 **A1:** build QSM from TLS.  
26  
27  
28 136 **A2:** extract  $U_d$  from QSM.  
29  
30 137  
31  
32  
33 138 Main cycle B:  
34  
35 139 **B1:** simulate SSM (with fixed random generator seed for reproducibility) for the given parameters and  
36  
37  
38 140 extract  $U_m$ .  
39  
40 141 **B2:** compare  $U_m$  and  $U_d$  getting an estimation of the distance  $D$  between them.  
41  
42  
43 142 **B3:** change SSM parameters trying to decrease  $D$ , go to B1 or stop and go to B4 (changing of the  
44  
45 143 parameters and stopping criteria depend on any particular realization of the optimization routine).  
46  
47  
48 144 **B4:** simulate SSM with the “best-fit” parameter values corresponding to the smallest found  $D$ .  
49  
50 145 **B5:** loose the randomness of the best-fit SSM and generate morphological clones.  
51  
52 146  
53  
54  
55 147 At the preparation stage, the QSM is formed from the TLS point cloud (A1). The detailed description  
56  
57 148 of this process is reported in [18, 19]. The resultant QSM contains all geometrical and topological  
58  
59  
60  
61  
62  
63  
64  
65

149 features needed to form the empirical distributions  $U_d$ . The distributions can be formed for several tree  
1  
2  
3 150 individuals if they are close by shape to ensure the sample size.

4  
5 151  
6  
7 152 At the main cycle of the algorithm, the empirical distribution  $U_m$  is formed from the simulated SSM  
8  
9  
10 153 tree (B1). Next,  $U_m$  is compared against  $U_d$  using the measure of distance (B2). The optimization  
11  
12 154 routine iteratively minimizes the distance value every time changing the parameter values of SSM  
13  
14  
15 155 (B3), simulating SSM, and repeating the cycle from B1. After the stopping criteria of the optimization  
16  
17 156 routine (number of iterations, minimal allowed distance etc.) are met, the algorithm stops and produces  
18  
19 157 the best-fit SSM tree (B4). The best-fit SSM with different random sequences produces different  
20  
21  
22 158 outcomes – morphological clones.

23  
24 159  
25  
26  
27 160 In Methods, we describe each of the main components of the algorithm in further detail.

28  
29 161  
30  
31  
32 162 **III. Testing of the algorithm**

33  
34 163  
35  
36 164 First, we run the optimization within each of the parameter groups  $I - V$ , representing different  
37  
38  
39 165 processes of growth (see Methods for details), to determine the basic values of the parameters. These  
40  
41 166 basic values represent choices that generate a viable tree structure with proportions and scale  
42  
43  
44 167 approximately equal to those of the target QSM. Each optimization run takes the best parameters for  
45  
46 168 the group optimized at the previous step. The target structural distributions  $U$  for these runs are  
47  
48  
49 169 segment-related ( $S$ ) features of the branches of topological order  $w = 0, 1$ , that is  $S^{0,1}$  (see the details of  
50  
51 170 the notations and description of the features in Methods). Note that this exercise serves a basic  
52  
53  
54 171 exploration of the model's behavior, which can be (partially) replaced, for example, by the expert  
55  
56 172 guesses for the parameter values or some calibration process (if the model is designed for specific  
57  
58 173 purposes and/or species).

174

1

2175 Second, based on these preliminary results we determine the most influential parameters for each of

3

4176 the group and combine them in a single optimization set up. Several independent optimization runs

6

7177 were taken in order to determine the most influential parameters. For example, we found that the

8

9178 angular properties vary the least among these runs, whereas the apical dominance requires subtler

11

12179 adjustments (as can be understood from the complex structure of the target QSM). This step is required

13

14180 to reduce optimization time, and it is not needed if one possesses large computational resources.

16

17181

18

19182 **Low order topological adjustment of the shape**

20

21183

23

24184 After these initial manipulations, we obtained a model with 11 parameters and good fit of the trunk

25

26185 ( $w = 0$ ) and first order branches (Fig. 2C) with classical metrics  $d_h = 0.05$ ,  $d_g = 0.42$ ,  $d_c = 0.57$  (see

28

29186 Methods for the definition of the classical metrics). However, the overall form of the resulting minimal

30

31187 score tree does not resemble the target QSM due to its rosette-shape (Fig. 2A, B). A closer look at the

33

34188 tree reveals that the higher order branches ( $w > 1$ ) are mainly responsible for the formation of the

35

36189 rosette-shape of the tree, i.e. the orders which were not subject to the optimization (Fig. 2). This

37

38190 example demonstrates the contribution of the higher order branches to the overall tree shape, which

40

41191 suggests using the scatters of these orders in further optimization steps. Moreover, the branch-related

42

43192 ( $B$ ) features, such as the angular properties of branches of order  $w > 1$ , were not captured well

45

46193 (Fig. 2E), although similar order segment-related  $S$ -features show right stochastic tendencies (Fig. 2D)

47

48194 generated automatically by the growth algorithm of the SSM. However, note that these features of

50

51195  $w > 1$  were not subject to optimization. This further stipulates usage of the branch-related  $B$ -scatters of

52

53196 orders  $w > 1$  (see the details of the notations and description of the features in Methods).

54

55197

57

58

59

60

61

62

63

64

65



198 **Figure 2: The rosette-shape SSM resulting from the adjustment of the low order segment-related**  
1  
2 199 **( $S^{0,1}$ ) scatters.** (A) The SSM tree; (B) the target QSM; (C) some segment related ( $S^{0,1}$ ) scatters used in  
3  
4  
5 200 the optimization; (D) higher order ( $w = 2$ )  $S$ -scatters (not used in optimization); (E) higher order ( $w =$   
6  
7 201 2, 3) branch related  $B$ -scatters (not used in optimization). SSM/QSM scatters are shown in red/blue.

8  
9  
10 202

## 11 12 203 **Low and high order topological adjustment**

13  
14  
15 204

16  
17 205 The increase in number of the structural feature tables is coupled with the increase in number of  
18  
19 206 distinct distance values, that is, each pair of tables (QSM vs. SSM) produces a distance score to be  
20  
21  
22 207 optimized. Although the optimization of the mean distance value for all tables hinders the  
23  
24 208 improvement for each table separately, the low order and high order branches need to be fitted to the  
25  
26  
27 209 corresponding branches of the target QSM as we have shown above (Fig. 2). To reduce the number of  
28  
29 210 distinct feature tables for the optimization we further utilize the merged data sets resulting in two joint  
30  
31  
32 211 segment- ( $S$ ) and branch-related ( $B$ ) tables for all topological orders (see Methods for description of the  
33  
34 212 merged data sets).

35  
36  
37 213

38  
39 214 Thus, we opted for  $S^{0,1}$  and  $B^{2,3,4}$  merged data sets in the next run of optimization to account for the  
40  
41 215 higher order branch variability (Fig. 3,  $d_h = 0.08$ ,  $d_g = 0.20$ ,  $d_c = 0.68$ ). No longer we can see the  
42  
43  
44 216 rosette-shape due to the correct account of the angular properties of the higher order ( $w > 1$ ) branches  
45  
46 217 (Fig. 3E). The poor convergence of the branch linear dimensions (radii, lengths etc.) present in the  
47  
48  
49 218 branch-related tables might be due to the parameter choice of the model. Namely, the small proportion  
50  
51 219 of branches demonstrating right  $R_f$  values (Fig. 3E) appears to be the result of the fixed segment length  
52  
53  
54 220 we opted for as a compromise between reality and computational complexity (the QSM minimal  
55  
56 221 segment length is close to zero, median is 0.06 m, whereas that of SSM is fixed to 0.2 m). Noteworthy  
57  
58 222 is the similar span of the curvature data points of SSM and QSM for  $w = 1, 2$  (Fig. 3C and D),  
59

60  
61  
62  
63  
64  
65

223 although  $w = 2$  branch curvature was not subject to the optimization. Additionally, due to the lack of  
1  
2224 the orientation landmark in the feature data sets our best-fit SSM is fitted to the target QSM with  
3  
4  
5225 accuracy of the rotation around Z-axis (this could be adjusted, for example, by associating South  
6  
7226 direction with a coordinate axis).

8  
9  
10227  
11  
12228 **Figure 3: Low and high order adjustment of the stochastic feature tables.** The best-fit SSM is  
13  
14  
15229 obtained through optimization against  $S^{0,1}$  and  $B^{2,3,4}$  merged feature data sets. (A) The best-fit SSM  
16  
17230 tree, (B) the target QSM tree, (C) some projection scatters from  $S^1$ , (D)  $S^2$  projection scatters, (E)  $B^2$   
18  
19231 and  $B^3$  projection scatters.

20  
21  
22232  
23  
24233 **Clonal nature of the best-fit SSM**

25  
26  
27234  
28  
29235 Due to the highly discrete and stochastic nature of the tree growth, the structural distance hyper-  
30  
31  
32236 surface in the space of the parameters is extremely abrupt (Fig. 4A). Hence, finding the global minima  
33  
34237 of such surface is not a trivial task (the classical smooth function optimizers are not suitable in this  
35  
36  
37238 case, while stochastic discrete optimizers, like the genetic algorithm, seem to be more appropriate).  
38  
39239 Moreover, the hyper-surface itself is a stochastic entity changing every time the new sample of random  
40  
41240 numbers is used for a particular SSM growth realization. Therefore, any best-fit SSM is the best for a  
42  
43  
44241 particular realization of this stochastic process: one needs to study variability of the tree shape and the  
45  
46242 chances are that other SSM growth realization can produce a lower distance value (Fig. 4B). We call  
47  
48  
49243 these many realizations of the SSM growth *morphological tree clones*.

50  
51244  
52  
53  
54245 **Figure 4: Stochastic structure distance profiles in the parameter space.** (A) Three realizations of  
55  
56246 the distance hyper-surface projection along a dimensionless parameter  $\lambda$  of the SSM, controlling the  
57  
58  
59247 apical dominance of a tree (the shown fragment of the projection with the step of 0.001 approximates

248 30% of the allowed variability of the parameter during optimization, which was [0.35, 0.65]). (B)  
1  
2249 Structural distance ( $U = \{S^{0,1}, B^{2,3,4}\}$ ) values for 100 randomly generated SSM trees for each value of a  
3  
4  
5250 discrete SSM parameter, i.e. number of growth iterations (red line connects the median points of the  
6  
7251 distance distributions for each parameter value; blue line shows the same median distance profile but  
8  
9  
10252 for the disturbed system, see (C)). (C) Same as in (B), but  $U = S^{0,1}$  (blue line is the median profile; red  
11  
12253 line is from (B)). The SSM is the best-fit SSM obtained in the experimentation reported in Fig. 3; the  
13  
14  
15254 black arrow indicates the parameter value of the best-fit SSM found in the experimentation.  
16  
17255  
18  
19256 The structural distance profile depends not only on the parameters of the SSM, but the choice of the  
20  
21  
22257 structural data sets. For example, in Fig. 4B and C the median distance profile is depicted given  $U =$   
23  
24258  $\{S^{0,1}, B^{2,3,4}\}$  (red line) and  $U = S^{0,1}$  (blue line). In the given parameter span the latter seems to be more  
25  
26  
27259 flattened and lifted compared to the former. The addition of the  $B^{2,3,4}$  data set might be seen as a  
28  
29260 perturbation to the distance profile changing the landscape properties (like minima). In our simulations  
30  
31  
32261 we maintain the global parameter boundaries, which allows for the search within the full available  
33  
34262 space. However, we sequentially improve the model characteristics by perturbing the system, i.e.  
35  
36  
37263 changing the parameters, their intervals, and the  $U$  data sets to address problematic parts of the SSM  
38  
39264 (like rosette-shape, Fig. 2) such that at every next optimization run the genetic algorithm is instructed  
40  
41265 to search around the previous best point using the initial ranges (see Methods for details).  
42  
43  
44266  
45  
46267 Given the considerations above about the nature of the structural distance hyper-surface, the further  
47  
48  
49268 study of the morphological clones is needed. Specifically, the variability and plausibility of the clonal  
50  
51269 shapes need to be addressed. For example, the clones must be further selected as to produce realistic  
52  
53  
54270 tree shapes (especially, when the general purpose SSM is used, like in this study), however, in our  
55  
56271 analysis we did not find any unrealistic tree sampled from the best-fit SSM (any specific application  
57  
58272 imposes additional constraints on the parameters, which results in removal of the unrealistic shapes).  
59  
60  
61  
62  
63  
64  
65

273 Additionally, the variability of the clones can be further calibrated, for instance, by the analysis of the  
1  
2274 natural/QSM clonal individuals.  
3

4  
5275  
6

7276 **Morphological tree clones**  
8

9  
10277  
11

12278 The quintessence of our work is the generation of the morphological clones. In our pipeline, this  
13  
14279 occupies the last stage (see Fig. 1, B5). After the optimization is finished and the best-fit SSM is  
15  
16  
17280 found, one can further randomize the outcome of SSM by letting the random number generator  
18  
19281 produce different sequences every time SSM is run. As a result, the different realizations of SSM  
20  
21  
22282 should constitute the morphological clone generator yielding structural copies close to QSM and to  
23  
24283 each other and *varying* in fine detail of organization of their branches. In other words, the coarse-grain  
25  
26  
27284 structure is repeated in each clone (and possibly grasps that of the target QSM), whereas the fine-grain  
28  
29285 structure varies.  
30

31  
32286  
33

34287 **Figure 5: Morphological clones generated from the best-fit SSM.** The best-fit SSM was found  
35  
36288 using the higher topological order adjustments (Fig. 3) with number of growth iterations 30 (A), 26  
37  
38  
39289 (B), and 18 (C). The height, girth, crown spread, and classical metrics distributions are shown in (D)  
40  
41290 for the clones in (A), (B), and (C) (the total number of generated clones for each case is  $n = 100$ , only 6  
42  
43  
44291 are shown). The black horizontal line indicates the corresponding measure of the target QSM.  
45

46292  
47

48  
49293 We demonstrate visualization of six clones for three distinct cases in Fig. 5 (clones from other best-fit  
50  
51294 SSM's are provided at [31]). One can see the fine-grain variation in the structure in each panel of the  
52  
53  
54295 figure, although the overall (coarse-grain) structure is preserved and presumably captures that of the  
55  
56296 target maple QSM (Fig. 6). The three models are: the one found during the optimization process (Fig.  
57  
58  
59  
60  
61  
62  
63

297 5A), the one minimizing the sample median distance profile for  $D_S(U = \{S^{0,1}, B^{2,3,4}\})$  shown in Fig. 4B  
1  
2298 (Fig. 5B) and one minimizing the sample median profile  $D_S(U = S^{0,1})$  from Fig. 4C (Fig. 5C).  
3  
4  
5299  
6  
7300 Out of 100 simulated clones for each case, we can see that the best-fit SSM obtained directly as the  
8  
9  
10301 optimization outcome (Fig. 5A) produces larger proportion of individual trees exhibiting the three  
11  
12302 standard allometric measures closer to those of QSM (Fig. 5D). However, we argue that such simple  
13  
14  
15303 description of a tree, as using the allometric measures, cannot be exhaustive enough to capture both the  
16  
17304 overall structure and its fine details. Moreover, such static measures are absolutely useless for  
18  
19305 generation of morphological clones.  
20  
21  
22306  
23  
24307 The height statistics have the largest variability but by the visual inspection of the drawn clones in  
25  
26  
27308 Fig. 5 one can see that this variability does not exert significant alterations of the Z axis span and the  
28  
29309 trees seem to have even heights. Perhaps, the way we calculate the height of a tree produces such large  
30  
31  
32310 deviations in each particular case, which makes it a non-robust estimator (see Methods for the details  
33  
34311 of the height calculation).  
35  
36312  
37  
38  
39313 Similarly, the girth estimation, although being captured decently, produces large errors  $d_g$ , which  
40  
41314 seems to be a result of variation in its linear dimensions (Fig. 5D). The girth dimension spans a small  
42  
43  
44315 proportion of the dimension of the whole tree: from several to tens of centimeters compared to meters  
45  
46316 of the whole tree. This makes the girth specific error look gigantic (exceeding in some cases 100%)  
47  
48  
49317 and thus non-robust as well.  
50  
51318  
52  
53319 The crown spread measure shows significant variation (Fig. 5D). We believe that this takes place due  
54  
55  
56320 to the environment of the real tree the QSM was reconstructed from, which was not modeled  
57  
58321 appropriately in the SSM. Namely, the environmental effects (positions relative to the sun, as the tree  
59  
60  
61  
62  
63  
64  
65

322 grows in the Northern country, animals, winds, neighboring trees etc.) might cause systematic  
1  
2323 influences exerted on the shape of the QSM tree. These influences were not accounted for in the SSM,  
3  
4  
5324 which was allowed to grow in any direction, limited by the uniform light conditions, existing branches  
6  
7325 of the same tree, and global boundaries of the available space. In addition to the environment  
8  
9  
10326 influences, there are TLS measurement and QSM reconstruction errors, arising from the physical  
11  
12327 limitations of the instrumental technique and stochasticity of the QSM formation, respectively (see  
13  
14  
15328 Methods).

16  
17329

18  
19330 Finally, the true understanding of the variability of any measures of the morphological clones comes  
20  
21  
22331 with the measurements of the real clones. Carrying out control experiments with QSM reconstructed  
23  
24332 from the real clonal individuals (with the application dependent definition of a clone, e.g. genetic  
25  
26  
27333 clones) can only assess the variability. These real clone controlled experiments can further identify  
28  
29334 whether the obtained variability is large/small for the given species/clones and lead to the adjustment  
30  
31  
32335 of the optimization parameters.

#### 33 34336

#### 35 36337 **IV. Bayes Forest toolbox**

#### 37

38  
39338

40  
41339 We have further developed a unified interface using Matlab, facilitating exploration, drawing,  
42  
43  
44340 optimization, and simulation of SSM and QSM as well as study of the morphological tree clones. Our  
45  
46341 interface allows for faster and easier manipulation of the required data, models, and optimization  
47  
48  
49342 routines from the Matlab Optimization Toolbox, using only the required elements of otherwise  
50  
51343 complex Matlab configuration for the analysis.

52  
53344

54  
55  
56345 The Bayes Forest toolbox is freely available at [31] (the version used in this study) and at [32] (the link  
57  
58346 is preferred for contributions and contains the latest version of the package). We also encourage the  
59  
60  
61  
62  
63  
64  
65

1 347 plant and computer scientists' community to expand their efforts using the toolbox with other species  
2  
3 348 and models. Such a systematic approach can further be useful in tinkering the best options for creating  
4  
5 349 QSM, SSM, and construction of the structural data sets.  
6

7 350

8  
9  
10 351 **V. Discussion**

11  
12 352  
13  
14 353 In this work, we described an algorithmic pipeline aimed at producing stochastic structural replicas, or  
15  
16  
17 354 morphological "clones", of trees from a QSM tree (data from TLS reconstruction) and a  
18  
19 355 complimentary SSM tree (analytical/procedural growth model). The pipeline is based on an iterative  
20  
21  
22 356 minimization of a distance between morphological structures. The distance is based on construction of  
23  
24 357 the structural data sets of the tree morphologies and subsequent measure of their discrepancy using the  
25  
26  
27 358 ideas of distribution tomography analysis. The resulting best-fit morphological clones are statistically  
28  
29 359 similar which is expressed in overall similarity of their form (coarse-grain), but, nevertheless,  
30  
31  
32 360 difference in fine details of structural organization (fine-grain).

33  
34 361

35  
36 362 Here, we have shown the general logic behind the pipeline and principle possibility for generation of  
37  
38  
39 363 the morphological clones as defined above. For this purpose we used a highly variable procedural tree  
40  
41 364 model [3], which is more difficult to optimize. As the pipeline consists of several elementary steps,  
42  
43  
44 365 each of which can be changed according to the application and target analysis, we have proposed an  
45  
46 366 initial set-up and basic configuration that are capable of the task we have set. We assume larger  
47  
48  
49 367 possibilities of exploration of the proposed configuration, let alone changing the steps and individual  
50  
51 368 algorithms within the pipeline, which could be fulfilled by the community of plant science researchers  
52  
53  
54 369 (for this reason, we also created a little toolbox in Matlab for easier representation and simulation of  
55  
56 370 the algorithm).

57  
58 371

372 Developing the principles of the pipeline, we were interested in biological plausibility of the results  
1  
2373 rather than visualization purposes. Thus, for example, we use real TLS measurements and general-  
3  
4  
5374 purpose measure of the distance, while omitting visual effects (e.g. shades, leaves etc.). We believe  
6  
7375 this pipeline can be useful in the rigorous analysis of the plant morphogenesis and corresponding  
8  
9  
10376 applications (in contrast to some similar studies done in computer graphics field, e.g. [5]).

11  
12377  
13  
14  
15378 Moreover, in our algorithm we employ the distance measure taking into account significant portion of  
16  
17379 the data (in fact, all data points of a given topological order(s)), not merely scalar overall entities  
18  
19380 proposed by other authors (e.g. [5, 33]). This allows for a more comprehensive analysis of forms and  
20  
21  
22381 their description, stemming from the statistical inference theory and in the spirit of Systems Biology  
23  
24382 studies. Due to this reason, we do not rely on the traditional metrics comparison in this work as we  
25  
26  
27383 found that similar values for the height, girth, and crown distances may correspond to different tree  
28  
29384 forms and, thus, be non-robust.

30  
31385  
32  
33  
34386 Use of several QSM trees can enhance the robustness of the statistical analysis presented here. In this  
35  
36387 case, similarly looking trees should be used and the degree of similarity might be established using our  
37  
38  
39388 definition of the structural distance. For example, the trunk features are more reliably reproduced in  
40  
41389 statistical sense, when several QSM's are used. In these lines, it might be stressed that other notions of  
42  
43  
44390 "clone" can be used to establish relationship with morphology. Thus, the genetic clones might be  
45  
46391 utilized to establish to what degree the morphology of a tree is encoded into genes (nature vs. nurture  
47  
48  
49392 problem).

50  
51393

## 52 53 54394 **Methods**

### 55 56 57395 58 59396 **I. Quantitative Structure Model (QSM)**



397

1

2398 QSM is derived from the point cloud obtained by TLS. Essentially, QSM is a surface reconstruction of

3

4399 the branches of the real tree measured by TLS. The reconstruction itself is a stochastic process, giving

6

7400 different architecture results for different runs. Therefore, the reconstruction introduces internal errors

8

9401 in addition to the TLS measurement errors. Besides giving spatial locations of parts of the tree, QSM

10

12402 also reconstructs topological relations between the tree branches. The branches of QSM consist of

13

14403 elementary units, i.e. circular cylinders, but other geometrical primitives can also be applicable [34].

15

16404 Thus, any potential structural information about the original tree can be approximated with high

18

19405 accuracy with QSM (details of the reconstruction algorithm are presented in [18, 19], for the validation

20

22406 of the algorithm see [19, 23-25]).

23

24407

25

26408 In this work, we use the reconstructed QSM of a maple tree (Fig. 6). The tree was measured in leaf-off

28

29409 conditions and our system consisted of a phase-based terrestrial laser scanner (Leica HDS6100 with a

30

31410 650–690 nm wavelength). The distance measurement accuracy and the point separation angle of the

32

33411 scanner were about 2–3 mm and 0.036 degrees, respectively. The horizontal distance of the scanner to

35

36412 the trunk was about 7–12 m, thus the average point density on the surface of the trunk (at the level of

37

38413 the scanner) for a single scan is about 2–5 points per square centimeter. The QSM of the subject maple

39

41414 tree consists of 19,000 cylinders approximating 3,078 branches.

42

43415

44

46416 **Figure 6: The target QSM structure in three main 2D projections.**

47

48

49417

50

51418 The subject QSM was chosen due to its non-trivial form and obvious irregularities in the tree growth.

53

54419 This is needed to determine whether the stochastic rules of SSM growth can account for this variability

55

56420 (which, in fact, might come from some deterministic sources, like constant wind, shading from the

57

58421 neighbors, animal influences etc., and which we do not know as we do not know the history of

60

61

62

63

64

65

422 growth). Thus, our algorithm tries to compensate the unknowns of the growth with simple stochastic  
1  
2423 rules of SSM and optimization of the stochastic distance function.

3  
4  
5424  
6

## 7425 **II. Stochastic Structure Model (SSM)**

8  
9  
10426

11  
12427 SSM is a simulated model, preferably based on analytical and/or heuristic rules for the tree growth;  
13  
14428 however, any viable algorithm for generating tree forms may be used. Importantly, the ultimate output  
15  
16  
17429 of the SSM simulation is a table containing data sets  $U$  (see IV.3 Structural data sets), describing the  
18  
19430 tree structure.

20  
21  
22431

23  
24432 Additionally, SSM may be supplied with stochastic variability in its parameter values. Through our  
25  
26  
27433 studies we implement simple stochastic variations (in the form of normal and uniform distributions)  
28  
29434 added to the parameter values of SSM.

30  
31  
32435

33  
34436 Finally, the elementary units forming the SSM branches should be similar to that of QSM for the  
35  
36437 appropriate comparison or, otherwise, any differences in the form primitives must be taken into  
37  
38  
39438 account. Usually cylinders are used in SSM studies and they were also shown, when used in QSM, to  
40  
41439 produce most reliable estimation of the real tree characteristics [34].

42  
43  
44440

45  
46441 Examples of SSM are: *LIGNUM* [13] – a functional-structural plant model based on the physiological  
47  
48  
49442 principles of growth of pine trees, but also applicable to other tree forms [35]; *self-organizing tree*  
50  
51443 *model* [3] is based on the heuristic principles of growth, the algorithm is capable of producing various  
52  
53  
54444 tree shapes and is used in computer graphics; *plastic trees* [4] are procedural growth models used in  
55  
56445 computer graphics; *AMAP/GreenLab* (see e.g. [36, 37]) is a modeling approach to generate FSPM  
57  
58446 based upon empirical rules of growth with some physiological processes taken into account.

59  
60  
61  
62  
63  
64  
65

447

1

2448 In this work, we use self-organizing tree model (SOT) with shadow propagation algorithm [3] as SSM

3

4449 with the minimal changes as to calculate the morphological features and produce the resulting data sets

6

7450 for comparison with QSM (in this work we used SOT implemented in the LPFG simulator, part of the

8

9451 Virtual Laboratory software suite [38], version 4.4.0-2424 for 64-bit Mac OS, see [39]). This

10

12452 procedural tree model is fast and able to generate variety of forms, hence we can use it effectively to

13

14453 optimize the whole algorithm in respect to technical details as well as to cover various tree shapes.

15

17454 Note that more specialized tree growth models designed for the species in question would be easier

18

19455 subjects for the morphology optimization, but, nevertheless, can be more valuable in biologically

20

22456 motivated studies (the usual choice is FSPM's, e.g. [30]).

23

24457

25

27458 The total number of growth parameters of the model is 27: 23 are grouped, 4 are fixed for all times.

28

29459 The values of the latter are dictated both by suggestions of the authors in [3] and the compromise

30

31460 between computation time and details of the morphological description. For example, the segment

32

34461 length is 0.2 m (we found this optimal to grow a full size tree within a reasonable span of time,

35

36462 although this is not the minimum length of the target QSM segments), the voxel size is 0.2 m, and the

37

39463 model tree grows within 12x12x12 m cube from the center of XY plane of the cube (Z-axis is oriented

40

41464 upwards).

42

43465

44

46466 The grouped parameters are divided between 5 distinct groups corresponding to different related

47

48467 processes:

49

51468 *Group I*: the initial growth parameters, including limiting values, and pipe model related parameters.

52

53469 *Group II*: environmental effects such as sensing of the neighborhood shading, vertical gradient of the

54

56470 light, tropism etc.

55

58471 *Group III*: apical dominance parameters.

59

60

61

62

63

64

65

472 *Group IV*: shadow propagation related constants (see [3]).

1

2473 *Group V*: angular/branching properties.

3

4  
5474

6

7475 **III. Structural data sets (U)**

8

9  
10476

11

12477 Structural data sets for any given tree structure are empirical collections of the physical dimensions

13

14478 and spatial orientation measures of segments and branches that are composed of segments. These data

15

16479 sets must be similarly obtained for any pair of  $\{U_m, U_d\}$  that is to be compared by means of the distance

17

18480 algorithm.

19

20  
21481

22

23482 Quantities in the data sets may represent scalar characteristics and/or relations between several

24

25483 covariates (e.g. radii, lengths, angles, tapering function of a branch etc.). On the one hand, one needs to

26

27484 exhaustively describe morphology of the tree using various geometrical and topological features. On

28

29485 the other hand, as the number of compared data sets  $\{U_m, U_d\}$  grows the efficiency of the optimization

30

31486 routine decreases, since the number of distance measures to be minimized grows correspondingly (one

32

33487 distance value for each pair  $\{U_m, U_d\}$ ). Thus, one needs more compact representation of the data. One

34

35488 solution is to use larger data sets with all possibly needed (for a given application) features. (Another

36

37489 solution is to use multi-objective optimization routines finding, e.g. Pareto front, though we do not

38

39490 employ such an approach in this work.) Therefore, we use larger tables of all measured features; hence,

40

41491 one table represents a data set. However, we are unable to merge segment- and branch-related features

42

43492 into a single table as these differ in dimension (Table 1). Thus, we usually compare the array of pairs

44

45493  $\{U_m, U_d\}$ , having as a result the array of distance values, but with such larger table representation we

46

47494 have smaller size of these arrays.

48

49  
50495

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

496 Branch- and segment-related data are described in Table 1 and Fig. 7. Throughout the manuscript we  
 1  
 2497 maintain the notations  $B^w$  and  $S^w$  for the branch and segment-related data sets of the (Gravelius) order  
 3  
 4  
 5498  $w$ , respectively. The zero order  $w$  is assigned to the trunk (a branch connecting the tree with the  
 6  
 7499 ground). At the branching points, the lateral buds give rise to branches with order  $w+1$ , where  $w$  is the  
 8  
 9  
 10500 order of the parent branch, while the apical buds continue the branch of the same order.

11  
 12501

14502 **Table 1: Branch and segment features.**  
 15

<b>Branch features, units</b>	Description
$\beta$ , degree	Inclination angle of the branch, i.e. angle with its parent branch.
$\alpha$ , degree	Azimuthal angle of the branch, i.e. angle around its parent branch (calculated from the fixed direction).
$L_t$ , m	Total length of the branch (calculated as the sum of the segment lengths constituting the branch).
$R_f$ , m	Initial radius of the branch, i.e. radius of its first segment.
$L_a$ , m	Length of over the parent branch from its beginning segment to the point where the current (child) branch emanates.
<b>Segment features, units</b>	Description
$R$ , m	Radius of the segment.
$L$ , m	Distance from the beginning of the branch to the segment.
$\gamma$ , degree	Angle between horizontal projections of the segment and its parent.
$\zeta$ , degree	Angle between vertical projections of the segment and its parent.

55503

56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65

504 **Figure 7: Visual structure of a tree and its representation using the structural data sets  $U$ .** (A) A  
1  
2505 sample tree; (B) geometrical features of the branch- ( $B$ ) and segment-related ( $S$ ) data sets; and (C)  
3  
4  
5506 various projections of the  $U$  data sets.  
6  
7  
8507 These features are not exhaustive and can be augmented at will, but we found this set sufficient for  
9  
10508 obtaining realistic tree shape outcomes. Representation of the data sets in the form of large branch and  
11  
12  
13509 segment related tables reduces the complexity of optimization process by reducing the number of  
14  
15510 distance values to minimize. Additionally, such representation of the data allows for the fast extraction  
16  
17  
18511 of all required relations between covariates or scalar entities without having them as separate data sets.  
19  
20512  
21  
22513 In a simulated SSM structure the extraction of topological relations between branches is  
23  
24  
25514 straightforward as the user observes the whole process of growth: the lateral buds start the next order  
26  
27515 and apical buds continue the current order. However, this is not the case with QSM since it is a time  
28  
29  
30516 snapshot of a tree form that does not retain the history of the tree growth. Thus, the reconstruction  
31  
32517 algorithm requires other principles for extraction of topology. Although the reconstruction algorithm  
33  
34  
35518 defines a complicated procedure that outlines the topology of a tree, it could be roughly approximated  
36  
37519 by the following rule: at branching points the thickest branch is the continuation of the same order  $w$ ,  
38  
39  
40520 while thinner branches are lateral expansions of the order  $w + 1$  [18]. For the species with weak apical  
41  
42521 dominance (shrubby trees) we maintain similar procedure when simulating corresponding SSM (for  
43  
44522 the species with strong apical dominance, both techniques should converge to the same result).  
45  
46  
47523  
48  
49524 Finally, it is possible to merge the corresponding data sets, which results at maximum in two large data  
50  
51  
52525 sets of branch- and segment-related features, respectively. While this simplifies the search of the  
53  
54526 distance minimum (max two values to minimize), this technique must be used with care as in this case  
55  
56  
57527 one heavily relies upon the growth rules of SSM. If these rules are not based on biologically motivated  
58  
59528 rules, SSM can produce highly unrealistic tree forms as the “best-fit”, since there is a possibility to mix  
60  
61  
62  
63  
64  
65

529 the features of different topological orders. For example, the branches of higher order could be much  
1  
2530 thicker than those of the lower order, which is unrealistic and naturally is taken care of in the  
3  
4  
5531 biologically based growth algorithms (e.g. pipe model).  
6

7532  
8  
9  
10533 **IV. Measure of structural distance ( $D_S$ )**  
11

12534  
13  
14535 The distance  $D_S$  between any two data sets, or empirical distributions (dimension or number of  
15  
16  
17536 variables of which is not limited), measures the difference between the local densities of the points in  
18  
19537  $U$ -space for these data sets (i.e. large segment- ( $S$ ) or branch-related ( $B$ ) tables of morphological  
20  
21  
22538 features). Here, it is constructed by measuring SSM vs. QSM difference of the normalized cumulative  
23  
24539 distributions of the point densities projected onto a number of line directions in the coordinate space of  
25  
26  
27540 the variables in  $U$ . The directions of lines are generated with quasi-Monte Carlo method using low-  
28  
29541 discrepancy (quasi-/sub-random) sequences, which cover the given space more evenly than uniformly  
30  
31  
32542 generated sequences. The difference between the projected cumulative distributions is further  
33  
34543 measured by the Kolmogorov-Smirnov statistic (any other can be used) and the resulting distance  
35  
36544 between the two data sets  $U$  is an average of all statistics calculated from each of the lines (see  
37  
38  
39545 Fig. 8A).  
40

41546  
42  
43  
44547 In general,  $U \in R^N$ , in our case  $N = 4$  (segment) or  $N = 5$  (branch) as can be seen from Table 1. The  
45  
46548 empirical probability density function  $p(U)$  can be approximated by the series of 1D density functions  
47  
48  
49549  $p_{1D}(U, L)$ , where  $L$  is a line in  $R^N$ , each of these 1D functions is constructed by projecting all the data  
50  
51550 points of  $U$  (thus, it is not a marginal distribution) onto a line  $L$  (in total we use 1000 such line  
52  
53  
54551 directions formed quasi-randomly). Cumulative distributions  $P_{1D}(U_m, L_i)$  and  $P_{1D}(U_d, L_i)$  for each line  
55  
56552 direction  $L_i$  are compared, thus, for any given data set pair  $\{U_m, U_d\}$  the resultant distance value is:  
57  
58  
59  
60  
61  
62  
63  
64  
65

$$D_S(U_m, U_d) = \frac{1}{n} \sum_{i=1}^n K[P_{1D}(U_m, L_i), P_{1D}(U_d, L_i)],$$

1553

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

554 where  $n$  is the number of lines and operator  $K[\cdot, \cdot]$  returns the Kolmogorov-Smirnov statistic for the  
 555 given pair of 1D empirical cumulative distributions.

556 **Figure 8: Distribution tomography of the structural data sets (A) and classical metric for the**

557 **crown spread (B).** (A) Data points in  $U$  (projected here for simplicity onto  $(u_i, u_j)$  plane, i.e. in 2D) are  
 558 used to construct the projection onto a line  $L$ . Cumulative empirical distribution is calculated along  $L$   
 559 (red). Only one line is shown, although typically one should use sufficiently enough number of lines  
 560 (uniformly distributed over all directions) to describe the form of the distribution. (B) Top view of a  
 561 tree: spokes (red) emanate from the ground segment (green) extending up to the most distant points  
 562 (blue).

563 **Traditional metrics ( $d_x$ ).** In order to provide a reference to traditional tree measurement systems, we  
 564 also calculate three main tree characteristics that are used for describing a tree shape (Frank, 2010).  
 565 *Height* is calculated as the highest point of a tree. *Girth* is calculated as the diameter of the ground  
 566 segment (the breast-height diameter is not appropriate for the shrubby trees). *Crown spread* is  
 567 calculated as follows. First, on XY-plane (top view, Fig. 8B) the set of spokes (red lines in Fig. 8B)  
 568 emanating from the center of a tree (the ground segment, green circle) is formed (here, we opted for  
 569 the spokes with azimuthal separation of 10 degrees). Then the length of each spoke is calculated as a  
 570 distance from the tree center to the most distant point of the crown in the direction of the spoke (blue  
 571 circles). The crown spread is twice an average of all spokes of a tree.

572 Finally, when comparing two tree shapes we calculate the distances as follows:



$$d_h = \frac{|h_d - h_m|}{h_d}; d_g = \frac{|g_d - g_m|}{g_d}; d_c = \frac{|c_d - c_m|}{c_d}.$$

1576  
2  
3  
4577 In this,  $h_d$ ,  $g_d$ , and  $c_d$  are the height, girth, and crown spread of the QSM tree, respectively, whereas  $h_m$ ,  
5  
6578  $g_m$ , and  $c_m$  are the corresponding entities of the best-fit SSM tree. Thus, the classical distance  $d_x$  shows  
7  
8  
9579 how large is the difference between entities  $x$  in proportion of the corresponding reference/QSM tree  
10  
11580 value.

12  
13  
14581

## 15 16582 V. Optimization routine

17  
18583  
19  
20

21584 The measure of structural distance  $D_S(U_m, U_d)$  is minimized by adjusting the parameters  $v$  of SSM.

22  
23585 In principle (with infinite sampling),  $D_S = 0$  for two trees (or, more precisely, infinitely large groups of  
24  
25  
26586 stochastically varying trees) that have exactly the same parameters  $v$ . These trees are not copies of each  
27  
28587 other, but they are structurally (statistically) similar. The choice of the  $U$  defining  $D_S$  is not unique, but  
29  
30  
31588 ideally well-chosen  $U$  should satisfy the following uniqueness condition for  $D_S$  to yield an acceptable  
32  
33589 measure of distance. Let three trees be given by  $v_A$ ,  $v_B$ , and  $v_C$ . Then, if  $D_S(U_A, U_B) < D_S(U_A, U_C)$ , one  
34  
35  
36590 can update  $C \leftarrow B$ , find any new  $v_B$  for which the inequality holds, and repeat until  $D_S(U_A, U_B) \rightarrow 0$  and  
37  
38591  $v_B \rightarrow v_A$ . In practice, this should be true in a large enough neighborhood of  $v_A$  (any steps down the right  
39  
40  
41592 valley lead to its bottom); however,  $D_S > 0$  due to the finite sampling and insufficient model.

42  
43593  
44

45  
46594 Any algorithm from a standard optimization library (e.g. Matlab Optimization Toolbox) that finds a  
47  
48595 minimum of an objective function ( $D_S = F(v)$ ) can be used. However, to facilitate global minimum  
49  
50  
51596 search and given the nature of the problem we use the genetic algorithm (implemented in Matlab,  
52  
53597 version R2015b). Additionally, some parameters of SSM may take only integer values, so the genetic  
54  
55598 algorithm handles the integer parameters correctly unlike, for example, the classical steepest decent  
56  
57  
58599 algorithm. The genetic algorithm iteratively finds a minimum of  $D_S$ , each iteration being called  
59  
60600 *generation*. Each generation is characterized with a number of individuals, i.e. *population*; one

601 individual is equivalent to one set of the parameter values. The variation is controlled by the *crossover*  
1  
2602 *rate* (rate of recombination of the population parameters) and *mutation rate* (rate of introduction of the  
3  
4  
5603 new variability into the population). The former is fixed to 80% in the Matlab Optimization Toolbox,  
6  
7604 whereas the latter is controlled by our configuration. The user controls ranges of the parameters. There  
8  
9  
10605 are two types of ranges: *global* lower and upper boundaries for each of the parameter values and *initial*  
11  
12606 *range*, from which the algorithm tries to construct the initial population (and, perhaps, where the best  
13  
14  
15607 solution lies). The latter controls the convergence rate: if it is too broad poor convergence is attained.  
16  
17608 Finally, algorithm stops when there have passed a fixed number of generations without improving the  
18  
19609 distance.

20  
21  
22610  
23  
24611 Thus, the objective function takes the input parameters  $v$ , simulates SSM with  $v$ , calculates and returns  
25  
26  
27612 structural data sets  $U_m$ . Subsequently, the objective function calculates  $D_S(U_m, U_d)$  and returns it to the  
28  
29613 optimization routine. The SSM, being a stochastic model, *must* have a fixed random generator seed  
30  
31  
32614 during optimization, i.e. the same input parameter set must produce the same structural output. This is  
33  
34615 needed for convergence of the optimization. After obtaining the final best-fit form of SSM, one can  
35  
36616 further explore the variability coming from different random number sequences used in the SSM  
37  
38  
39617 simulations (in addition to Matlab, we used GNU Octave version 4.2.0 for clone generation, see [40]).  
40  
41618 Thus, such random best-fit SSM is capable of producing the clonal morphologies (the same overall  
42  
43  
44619 structure with varying details of organization), which is the main goal of our algorithm.

45  
46620

47

48

## 49621 **Availability of supporting source code and requirements**

50

51

52622 Project name: Bayes Forest

53

54623 Project home page: <http://math.tut.fi/inversegroup/app/bayesforest/v1/>

55

56

57624 Operating system: Platform independent

58

59625 Programming language: Matlab

60

61

62

63

64

65

626 Other requirements: VLAB software suite, version  $\geq$  4.4.0-2424

1

2627 License: MIT

3

4

5628

6

## 7629 **Data availability**

9

10630 All data needed to reproduce the results of this study, some additional materials, and Bayes Forest

11

12631 Toolbox are available at [31]. The most recent version of the Toolbox is also available at [32] (this

13

14632 interface is preferred for the contributors and also contains the most recent version of the software).

15

16633

17

## 18634 **List of Abbreviations**

19

20635 FSPM – functional-structural plant model.

21

22

23636 QSM – quantitative structure model.

24

25637 SSM – stochastic structure model.

26

27638 SOT – self-organizing tree model.

28

29639 TLS – terrestrial laser scanning.

30

31640

32

## 33641 **Ethics approval and consent to participate**

34

35642 Not applicable

36

37643

38

## 39644 **Consent for publication**

40

41645 Not applicable

42

43646

44

## 45647 **Competing interests**

46

47648 The authors declare that they have no competing interests

48

49

50649

51

## 52650 **Funding**

53

54

55

56

57

58

59

60

61

62

63

64

65

651 This work was supported by the Academy of Finland: Suomen Akatemia (Center of Excellence in  
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

652 Inverse Problem Research, one of the PI's is Mikko Kaasalainen).

654 **Author contributions**

655 IP performed all simulations, processed the data, and wrote the manuscript; MJ wrote the code for  
656 calculating the structural distance, discussed the results; MÅ contributed to Bayes Forest Toolbox; PR  
657 generated and provided for the QSM data, wrote the manuscript and discussed the results; MK  
658 conceived the study, discussed the results, and wrote the manuscript.

660 **Acknowledgments**

661 We would like to thank Risto Sievänen and Wojtek Palubicki for useful discussion and comments on  
662 the model design and implementation.

664 **References**

666 [1] Prusinkiewicz P. Modeling plant growth and development. *Current Opinion in Plant Biology*.  
667 2004;7:79-83.  
668 [2] Fourcaud T, Zhang X, Stokes A, Lambers H, Körner C. *Plant Growth Modelling and Applications:*  
669 *The Increasing Importance of Plant Architecture in Growth Models*. *Ann Bot*. 2008;101:1053-1063.  
670 [3] Palubicki W, Horel K, Longay S, Runions A, Lane B, Mech R, et al. Self-organizing tree models  
671 for image synthesis. *ACM Transactions on Graphics*. 2009;28:58.  
672 [4] Pirk S, Stava O, Kratt J, Abdul Massih Said M, Neubert B, Mech R, et al. *Plastic Trees: Interactive*  
673 *Self-Adapting Botanical Tree Models*. *ACM Transactions on Graphics*. 2012;31:50.  
674 [5] Stava O, Pirk S, Kratt J, Chen B, Mech R, Deussen O, et al. *Inverse Procedural Modelling of Trees*.  
675 *Computer Graphics Forum*. 2014;33:118-131.

676 [6] Hallé F, Oldeman R, Tomlinson P. Tropical trees and forests: An architectural analysis. Berlin:  
1  
2677 Springer; 1978.  
3  
4  
5678 [7] Sachs T, Novoplansky A. Tree form: architectural models do not suffice. *Israel J Plant Sci.*  
6  
7679 1995;43:203–212.  
8  
9  
10680 [8] Room P, Hanan J, Prusinkiewicz P. Virtual plants: new perspectives for ecologists, pathologists  
11  
12681 and agricultural scientists. *Trends Plant Sci.* 1996;1:33-38.  
13  
14  
15682 [9] Sievänen R, Nikinmaa E, Nygren P, Ozier-Lafontaine H, Perttunen J, Hakula H. Components of  
16  
17683 functional–structural tree models. *Ann Sci.* 2000;57:399-412.  
18  
19684 [10] Godin C, Hanan J, Kurth W, Lacoite A, Takenaka A, Prusinkiewicz P, et al., editors.  
20  
21  
22685 Proceedings of the 4th International Workshop on Functional–Structural Plant Models, June 7-11,  
23  
24686 Montpellier, France; 2004.  
25  
26  
27687 [11] Mäkelä A, Hari P. Stand growth model based on carbon uptake and allocation in individual trees.  
28  
29688 *Ecol Model.* 1986;33:204-229.  
30  
31  
32689 [12] Rauscher H, Isebrands J, Host G, Dickson R, Dickmann D, Crow T, et al. *ECOPHYS: An*  
33  
34690 *ecophysiological growth process model for juvenile poplar. Tree Physiol.* 1990;7:255-281.  
35  
36691 [13] Perttunen J, Sievänen R, Nikinmaa E, Salminen H, Saarenmaa H, Väkevä J. *LIGNUM: a tree*  
37  
38  
39692 *model based on simple structural units. Ann Bot.* 1996;77:87-98.  
40  
41693 [14] Lacoite A. Carbon allocation among tree organs: a review of basic processes and representation  
42  
43  
44694 *in functional–structural tree models. Ann For Sci.* 2000;57:521-533.  
45  
46695 [15] Rosell J, Llorens J, Sanz R, Arnó J, Ribes-Dasi M, Masip J, et al. Obtaining the three-dimensional  
47  
48  
49696 *structure of tree orchards from remote 2D terrestrial LIDAR scanning. Agric and For Meteorol.*  
50  
51697 2009;149:1505-1515.  
52  
53  
54698 [16] Van Leeuwen M, Nieuwenhuis M. Retrieval of forest structural parameters using lidar remote  
55  
56699 *sensing. Eur J For Res.* 2010;129:749–770.  
57  
58  
59  
60  
61  
62  
63  
64  
65

700 [17] Rutzinger M, Pratihast A, Oude Elberink S, Vosselman G. Detection and modelling of 3D trees  
1  
2701 from mobile laser scanning data. In: International Archives of Photogrammetry, Remote Sensing and  
3  
4702 Spatial Information Sciences. 2010;XXXVIII:520-525.  
5  
6  
7703 [18] Raumonon P, Kaasalainen M, Åkerblom M, Kaasalainen S, Kaartinen H, Vastaranta M, et al. Fast  
8  
9704 Automatic Precision Tree Models from Terrestrial Laser Scanner Data. Remote Sensing. 2013;5:491-  
10  
11705 520.  
12  
13  
14706 [19] Calders K, Newnham G, Burt A, Murphy S, Raumonon P, Herold M, et al. Nondestructive  
15  
16707 estimates of above-ground biomass using terrestrial laser scanning. Methods in Ecol Evol. 2015;6:198-  
17  
18708 208.  
19  
20  
21709 [20] Xu H, Gossett N, Chen B. Knowledge and Heuristic Based Modeling of Laser-Scanned Trees.  
22  
23710 ACM Transactions on Graphics. 2007;26:19.  
24  
25  
26711 [21] Livny Y, Yan F, Olson M, Chen B, Zhang H, El-Sana J. Automatic Reconstruction of Tree  
27  
28712 Skeletal Structures from Point Clouds. ACM Transactions on Graphics. 2010;29:151.  
29  
30  
31713 [22] Preuksakarn C, Boudon F, Ferraro P, Durand JB, Nikinmaa E, Godin C. Reconstructing Plant  
32  
33714 Architecture from 3D Laser scanner data. In: DeJong T., Da Silva D, editors. Proceedings of the 6th  
34  
35715 International Workshop on Functional-Structural Plant Models. 2010. 14-16.  
36  
37  
38716 [23] Hackenberg J, Spiecker H, Calders K, Disney M, Raumonon P. SimpleTree - an efficient open  
39  
40717 source tool to build tree models from TLS clouds. Forests. 2015;6:4245-4294.  
41  
42  
43718 [24] Kaasalainen S, Krooks A, Liski J, Raumonon P, Kaartinen H, Kaasalainen M, et al. Change  
44  
45719 Detection of Tree Biomass with Terrestrial Laser Scanning and Quantitative Structure Modeling.  
46  
47720 Remote Sensing. 2014;6:3906-3922.  
48  
49  
50  
51721 [25] Raumonon P, Casella E, Calders K, Murphy S, Åkerblom M, Kaasalainen M. Massive-scale Tree  
52  
53722 Modelling from TLS Data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial  
54  
55723 Information Sciences. 2015;II-3/W4:189-196.  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

- 724 [26] Smith A, Astrup R, Raumonen P, Liski J, Krooks A, Kaasalainen S, et al. Tree Root system  
1  
2725 characterization and volume estimation by terrestrial laser scanning. *Forests*. 2014;5:3274-3294.  
3  
4  
5726 [27] Kaasalainen M. Dynamical Tomography of Gravitationally Bound Systems. *Inverse Problems and*  
6  
7727 *Imaging*. 2008;2:527–546.  
8  
9  
10728 [28] Bracewell R. *Numerical Transforms*. Science. 1990;248:697-704.  
11  
12729 [29] Sievänen R, Perttunen J, Nikinmaa E, Kaitaniemi P. Toward extension of a single tree functional-  
13  
14730 structural model of Scots pine to stand level: effect of the canopy of randomly distributed, identical  
15  
16731 trees on development of tree structure. *Functional Plant Biology*. 2008;35:964–975.  
17  
18  
19732 [30] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M. Data-based stochastic  
20  
21733 modeling of tree growth and structure formation. *Silva Fennica*. 2016;50:1413.  
22  
23  
24734 [31] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M. Bayes Forest Toolbox.  
25  
26735 <http://math.tut.fi/inversegroup/app/bayesforest/v1/>. Accessed 16 Apr 2017.  
27  
28  
29736 [32] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M. Contribution interface to  
30  
31737 Bayes Forest Toolbox. <https://github.com/inuritdino/BayesForest>.  
32  
33  
34738 [33] Frank E. A Numerical Method of Plotting Tree Shapes. *Bull East Nat Tree Soc*. 2010;6:2-8.  
35  
36739 [34] Åkerblom M, Raumonen P, Kaasalainen M, Casella E. Analysis of Geometric Primitives in  
37  
38740 Quantitative Structure Models of Tree Stems. *Remote Sensing*. 2015;7:4581-4603.  
39  
40  
41741 [35] Lu M, Nygren P, Perttunen J, Pallardy S, Larsen D. Application of the functional-structural tree  
42  
43742 model LIGNUM to growth simulation of short-rotation eastern cottonwood. *Silva Fennica*.  
44  
45743 2011;45:431–474.  
46  
47  
48744 [36] De Reffye P, Fourcaud T, Blaise F, Barthelemy D, Houllier F. A functional model of tree growth  
49  
50745 and tree architecture. *Silva Fennica*. 1997;31:297-311.  
51  
52  
53746 [37] Yan H, Kang M, de Reffye P, Dingkuhn M. A Dynamic, Architectural Plant Model Simulating  
54  
55747 Resource-dependent Growth. *Annals of Botany*. 2004;93:591-602.  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 748 [38] Federl P, Prusinkiewicz P. Virtual Laboratory: an interactive software environment for computer  
2 graphics. In Proceedings of Computer Graphics International. 1999, p. 93-100.  
3  
4 750 [39] Prusinkiewicz P. Virtual Laboratory (VLAB) / L-studio homepage.  
5  
6 751 [http://algorithmicbotany.org/virtual\\_laboratory/](http://algorithmicbotany.org/virtual_laboratory/). Accessed 17 Apr 2017.  
7  
8  
9 752 [40] Eaton J. GNU Octave. <http://www.gnu.org/software/octave/doc/interpreter/>. Accessed 17 Apr  
10  
11 2017.  
12  
13

14 754  
15

16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



Figure 1

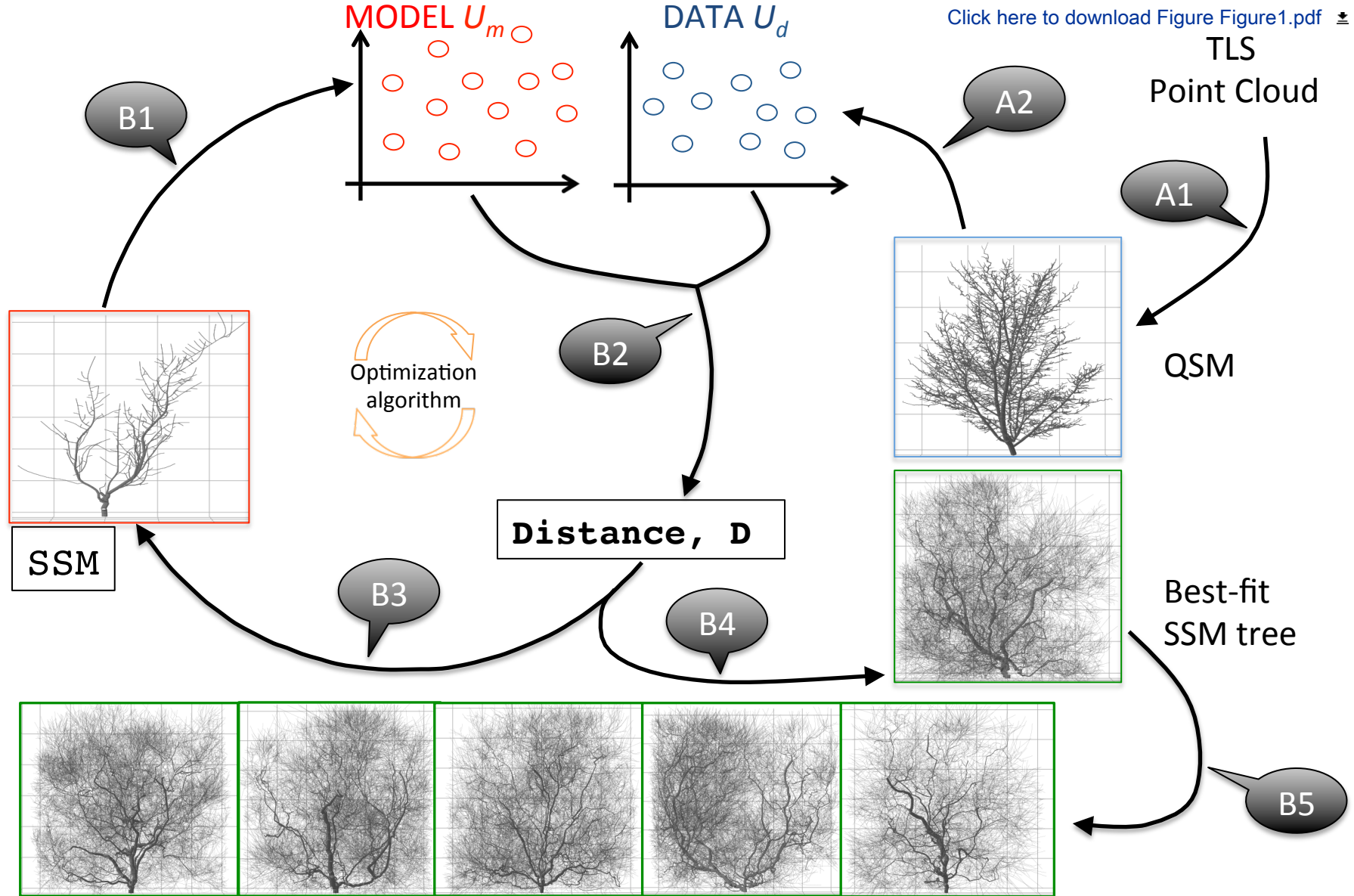


Figure 2

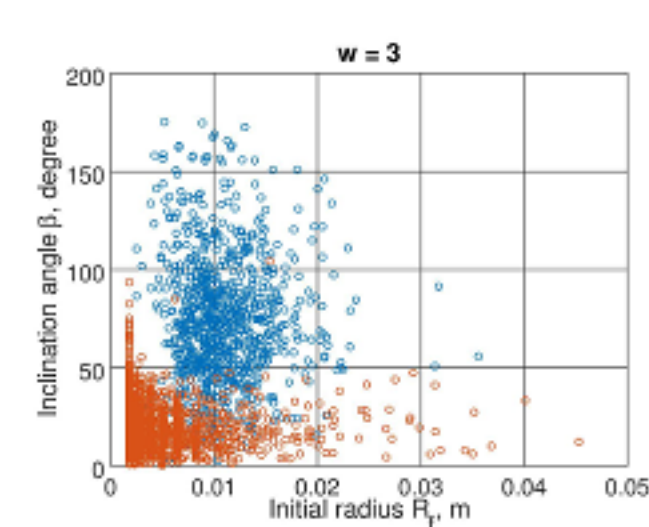
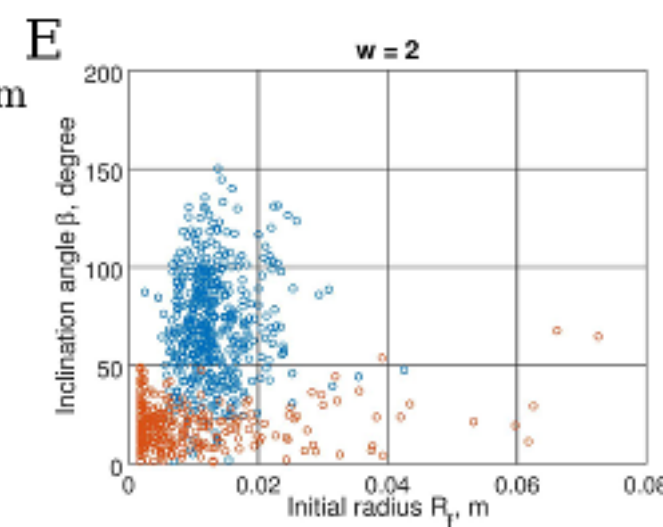
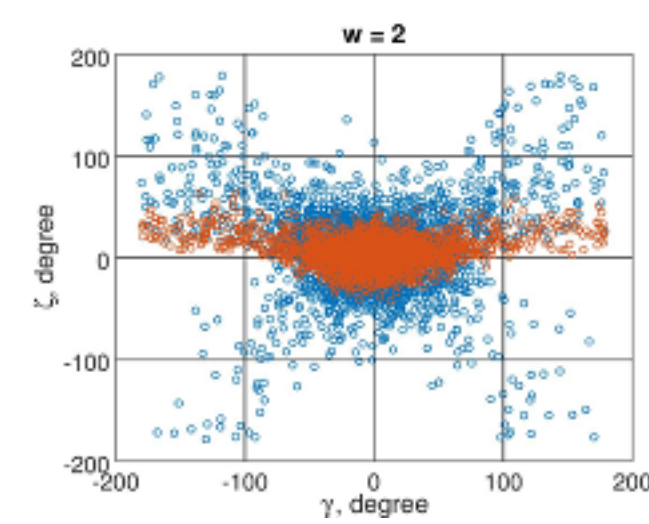
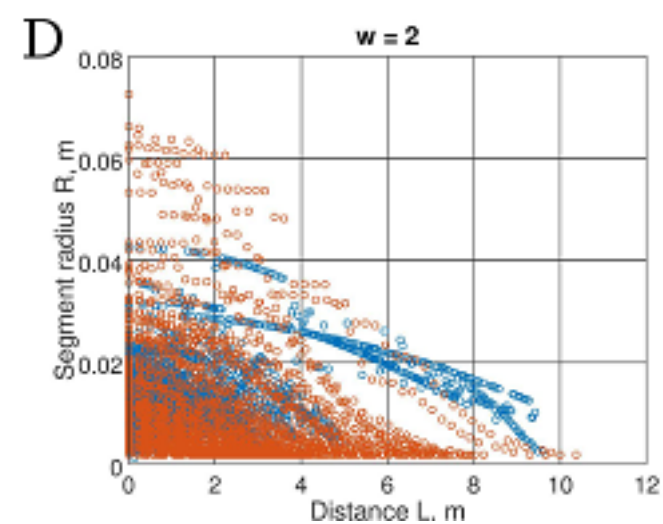
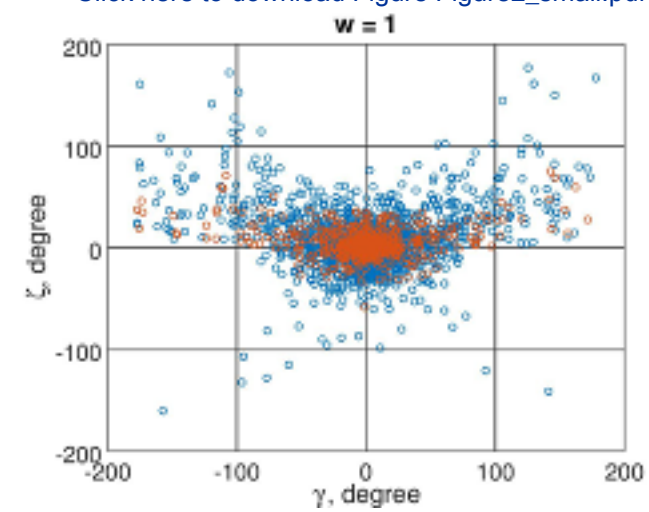
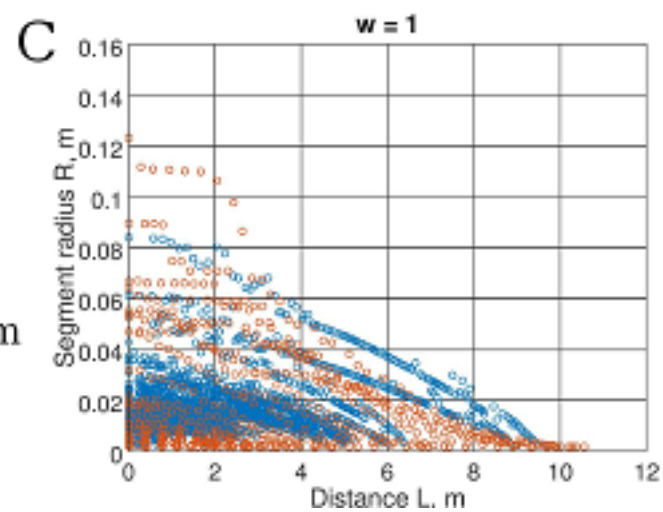
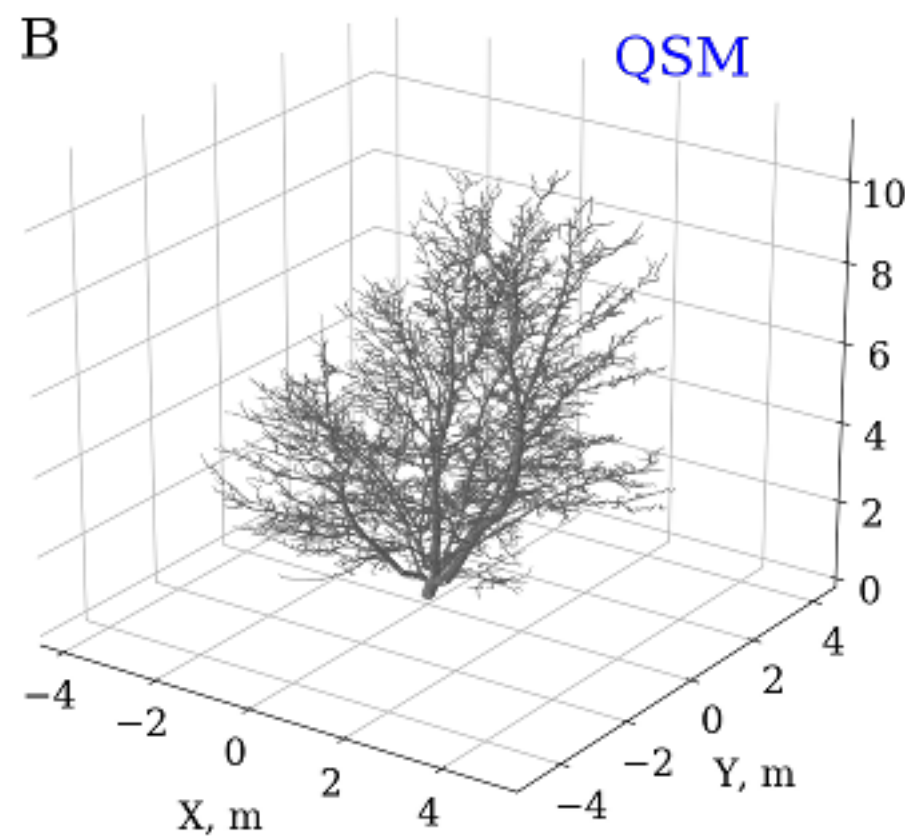
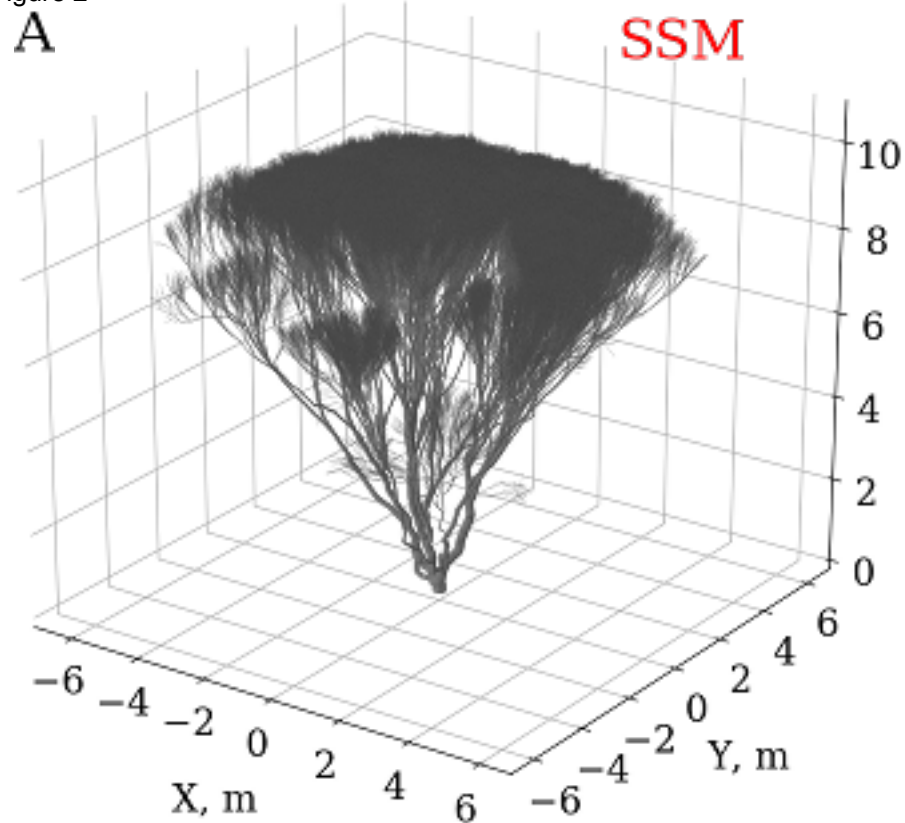
[Click here to download Figure2\\_small.pdf](#)

Figure 3

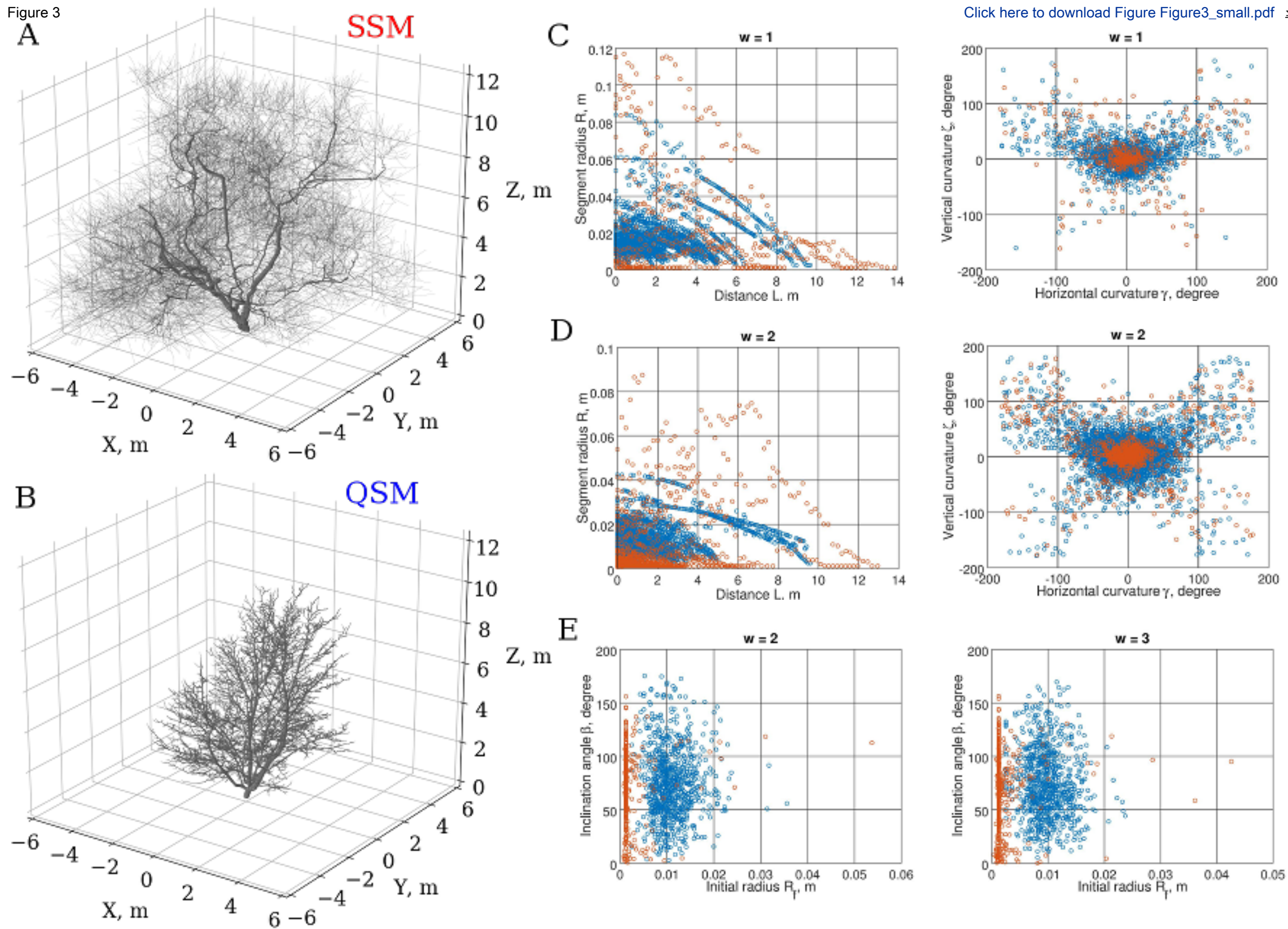
[Click here to download Figure3\\_small.pdf](#)

Figure 4

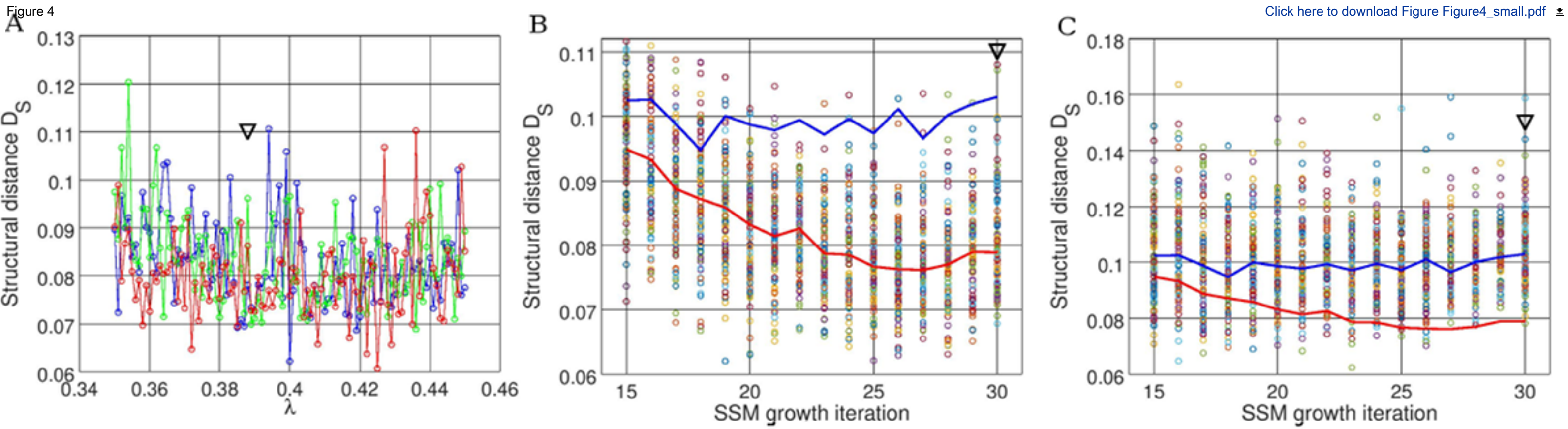
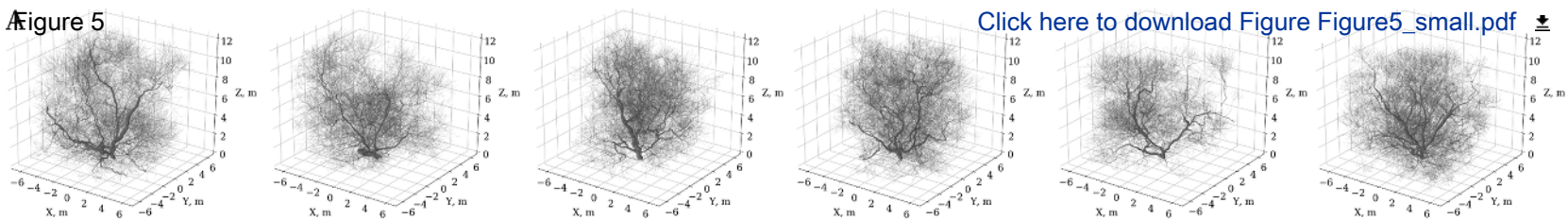
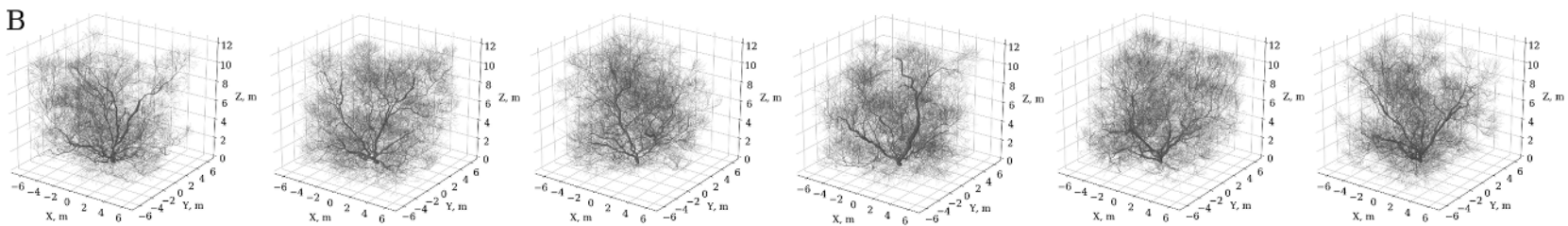


Figure 5

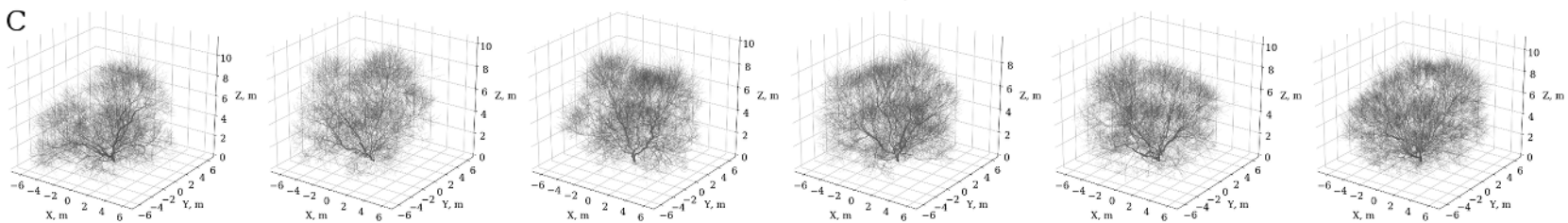


[Click here to download Figure5\\_small.pdf](#)

B



C



D

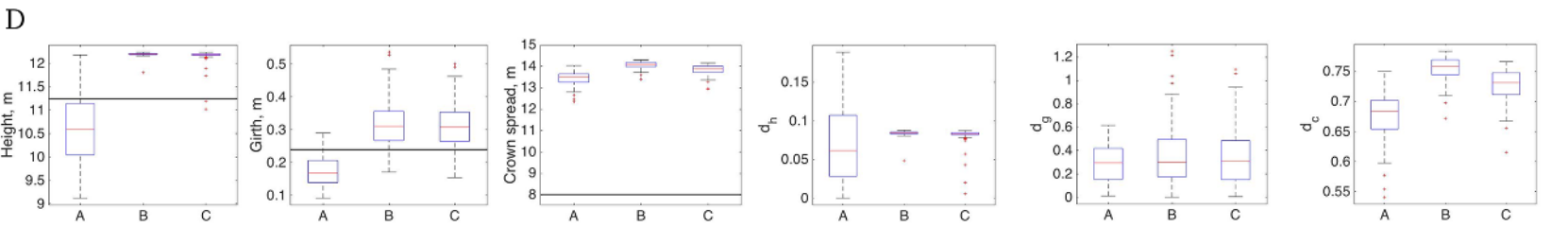
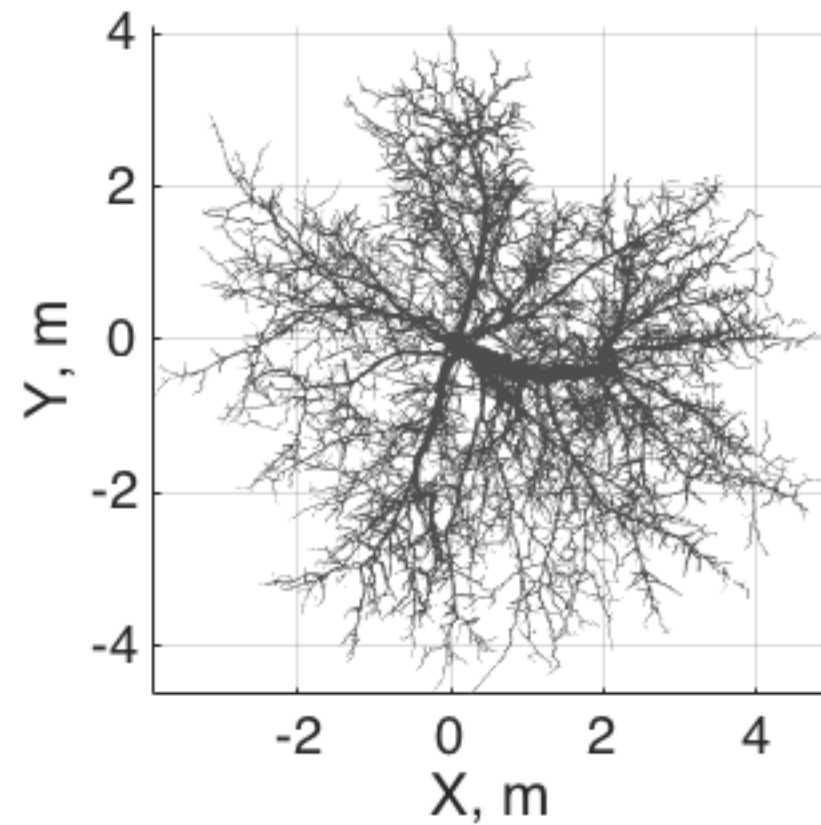
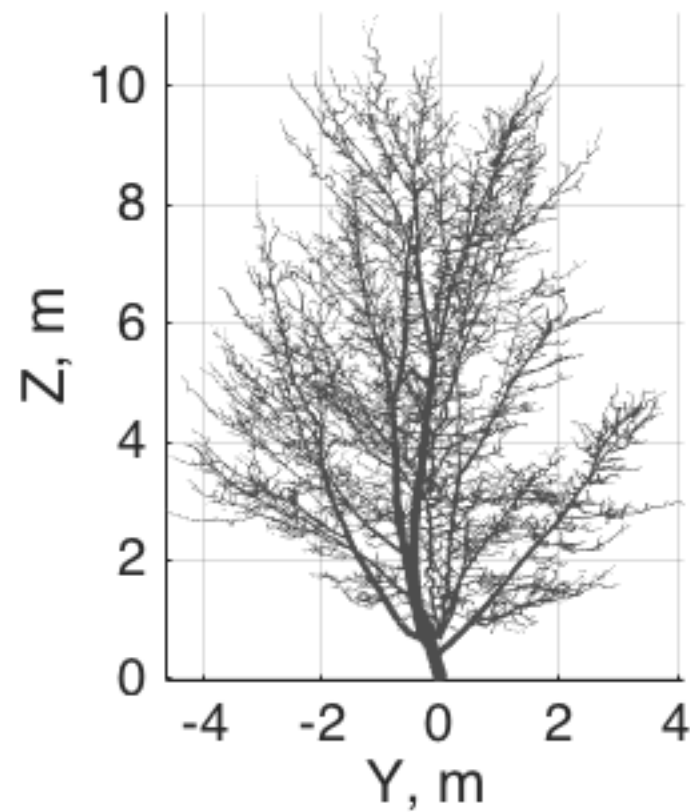
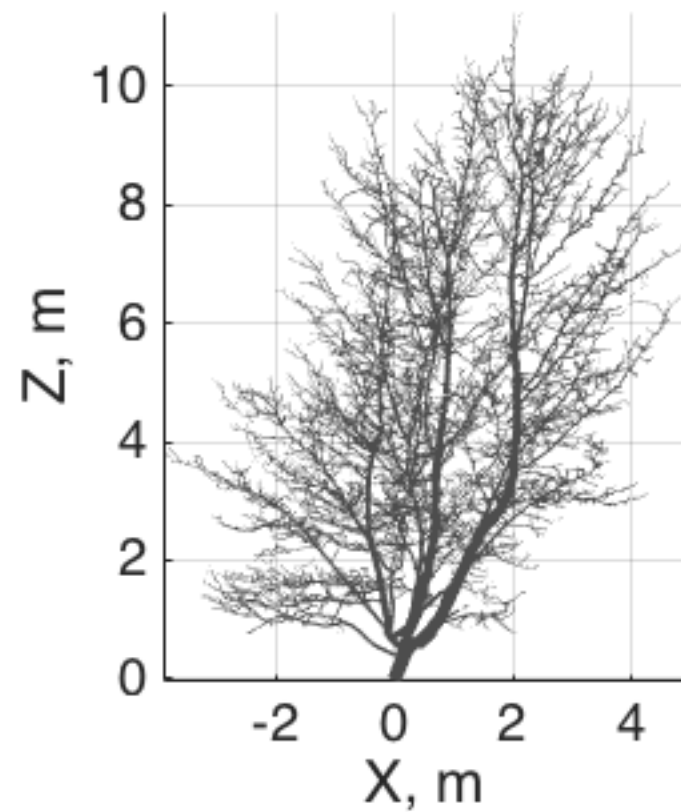
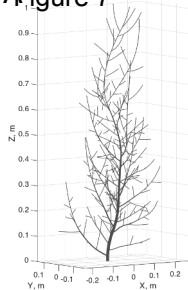


Figure 6

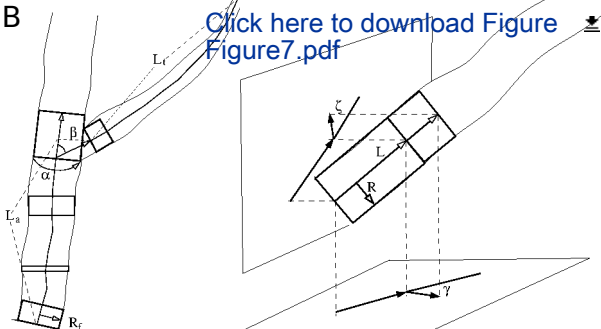
[Click here to download Figure Figure6.pdf](#)



**Figure 7**

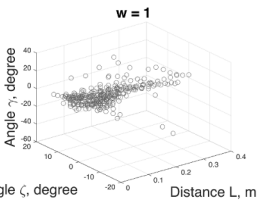
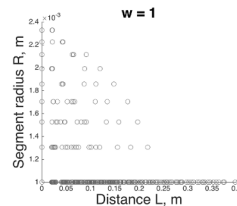
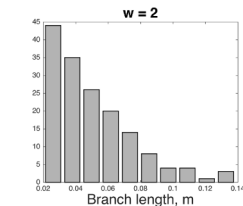
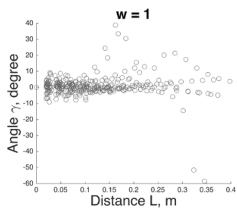
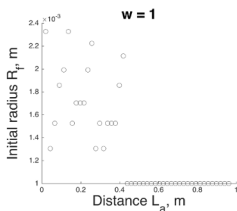
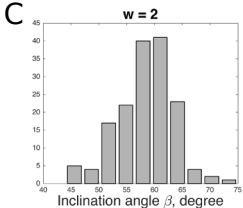


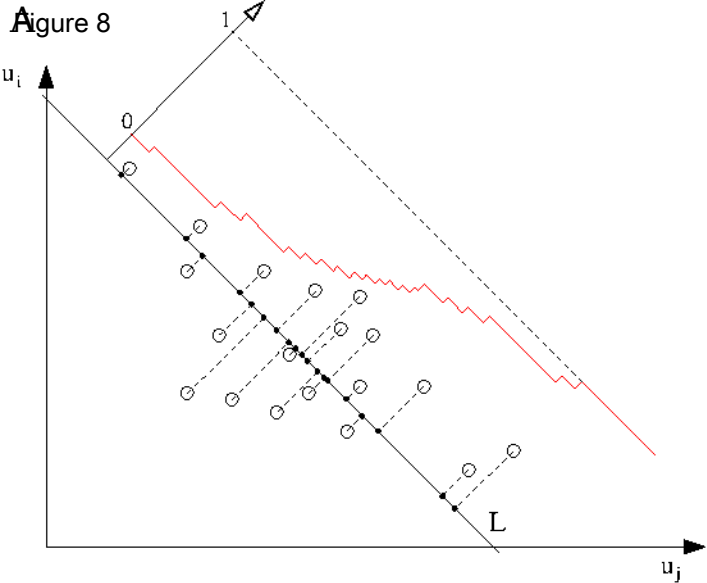
**B**



[Click here to download Figure Figure7.pdf](#)

**C**



**Figure 8****B** [Click here to download Figure Figure8.pdf](#) 