

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

Ilya Potapov¹, Marko Järvenpää², Markku Åkerblom¹, Pasi Raunonen¹, Mikko Kaasalainen¹

¹Mathematics Department, Tampere University of Technology, P.O. Box 553, 33101, Tampere,

Finland

²Helsinki 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.raunonen@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.

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 (and not identical) in respect of tree-level structure, but varying in fine-scale structural detail. 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 a 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 by means of empirical

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

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

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

34 **Findings**

36 **I. Background**

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

43 Modeling plant architecture is approached from many directions. Some progress has been achieved in
44 synthesis of realistic plant forms in the field of computer graphics [3-5]. These models, although based
45 on heuristic rules of growth, produce realistic tree representation in a fast and efficient manner, which
46 is usually dictated by the application of this approach, i.e. natural scene in computer visualization.
47 Heuristic growth rules of the procedural models for graphics applications are not firmly based on

1 48 biological principles, but nevertheless elucidate some algorithmic properties of the growth process (for
2 49 example, recursive [6] vs. self-organizing [3, 7] character of architecture development).
3

4
5 50

6
7 51 However, the most promising plant architectural models are so called functional-structural plant
8
9 52 models (FSPM) [8-10], because this type of models allows for a balanced description between
10
11 53 morphological and functional/physiological properties of a plant. It is capable of connecting the
12
13 54 external abiotic factors (e.g. radiation, temperature and soil) and the most vital functions of a plant
14
15 55 organism (such as photosynthesis, respiration, and water and salts uptake) with its structural
16
17 56 characteristics [1, 2].
18
19
20

21
22 57

23
24 58 Nevertheless, biologically relevant architectural plant models rely on data in a form of empirically
25
26 59 fitted functions and parameters that correspond to a particular species and/or certain site conditions
27
28 60 [11-14]. Thus, the change in these conditions requires re-calibration of the models, which is done in a
29
30 61 manual fashion every time the model is simulated for the new conditions. Strong dependence on data,
31
32 62 where each simulation would be calibrated automatically by data, is limited by both computation time
33
34 63 and lack of the fast measurement and processing systems allowing for a detailed 3D morphological
35
36 64 reconstruction of the actual plant/tree.
37
38
39

40
41 65

42
43 66 The most recent advances in laser scanning techniques allow for fast and non-destructive measurement
44
45 67 of trees with subsequent reconstruction of various characteristics depending on application (e.g. [15,
46
47 68 16]). Most of such studies dedicated to reconstruction of 3D point clouds obtained from laser scanning
48
49 69 measurements deal with overall characteristics, such as height, width, and volume of stems/crowns,
50
51 70 leaf index, biomass etc., resembling traditional destructive methods of measurement [15, 17].
52
53
54 71 However, the detailed precise geometrical and topological reconstruction of the tree architecture is
55
56 72 never achieved perfectly.
57
58
59
60
61
62
63
64
65

73

1
2 74 We use a fast, precise, automatic, and comprehensive reconstruction algorithm initially presented in
3
4
5 75 [18] and further developed and tested in [19]. The algorithm reliably reconstructs a quantitative
6
7 76 structure model (QSM), which contains all geometrical and topological characteristics of the object
8
9
10 77 tree. Input for the method is the 3D point cloud, sufficiently covering the tree, obtained from the
11
12 78 terrestrial laser scanning measurements (TLS). No additional allometric relations used for estimation
13
14
15 79 of the branch proportions (as in [20, 21]) are needed. Compared to other similar techniques (e.g. [20-
16
17 80 22]) this method requires few parameters and no user interaction. It reconstructs the tree surface with
18
19
20 81 subsequent cylinder (or any other geometrical primitive) approximation, which is usually consistent
21
22 82 with theoretical plant growth models. The reconstruction algorithm has been validated in several
23
24 83 studies with several different tree species and different scanner instruments [19, 23-26]. There are
25
26
27 84 other published QSM reconstruction methods from TLS data that can produce similar quality QSMs, at
28
29 85 least [23].
30

31 86

32
33
34 87 In this work, we utilize an inverse iterative procedure to optimize model's parameters for matching the
35
36
37 88 (empirical) distribution of structural features of the simulated stochastic tree models (FSPM, graphical
38
39 89 or other) to the tree reconstructed from the laser scanning data. Meanwhile, we formulate a measure of
40
41
42 90 similarity of the tree structures based on the tomographic analysis of the structural distributions (e.g.
43
44 91 Radon transform) [27, 28]. Finally, the optimal parameter set produces morphological "clone" trees
45
46 92 with similar overall structure, but with varying fine-scale details.
47

48 93

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

98 procedural model is easier to adapt for technical experimentation with the whole algorithm. Similar
1
2 99 algorithmic pipeline was reported in [5] for procedural tree growth models in the context of graphics
3
4
5 100 synthesis. However, in our approach we see the tree growth as a random process and, consequently,
6
7 101 apply corresponding statistical methods for measuring the similarity between trees. Moreover, in our
8
9
10 102 algorithm we put emphasis on biologically and physically relevant descriptions, hence, the careful
11
12 103 choice of the reconstruction algorithm. Another advantage is a possibility to use FSPM to relate
13
14
15 104 physiological parameters to the morphogenetic processes in trees. Finally, we use no extra structures
16
17 105 improving visual properties of trees but not supported by empirical observation (e.g. leaves). We note
18
19
20 106 that any other choices of parameters and feature descriptions can be used in our approach, further
21
22 107 facilitated with a programmable interface.
23

24 108

25

26

27 109 **II. Algorithm overview**

28

29 110

30

31

32 111 Our approach is based upon five distinct parts:

33

34 112

35

36

37 113

38

39 114

40

41

42 115

43

44 116

45

46 117

47

48

49 118

50

51 119

52

53

54 120

55

56

57

58

59

60

61

62

63

64

65

121 4. *Measure of structural dissimilarity, or structural distance D_S* , is a measure of discrepancy between
1
2 122 any two data sets. In other words, $D_S(U_1, U_2)$ returns a value quantifying how much different the
3
4
5 123 two data sets U_1 and U_2 are.
6

7 124 5. *Optimization algorithm* is a numerical routine capable of finding a minimum of any given function
8
9
10 125 by varying its arguments. Examples include Newton algorithm, genetic algorithm, and simulated
11
12 126 annealing.
13

14 127
15
16
17 128 The connection between these components is outlined in Fig. 1 with explanation in the text below.
18

19 129
20
21
22 130 **Figure 1: The algorithm outline (see explanation in the text).**
23

24
25 131
26
27
28 132 The algorithm outline (Fig. 1):
29

30 133
31
32
33 134 Stage A: Preparation
34

35 135 **A1:** build QSM from TLS.
36

37
38 136 **A2:** extract U_d from QSM.
39

40 137
41
42
43 138 Stage B: Main cycle
44

45 139 **B1:** simulate SSM (with fixed random generator seed for reproducibility) for the given parameters and
46
47 140 extract U_m .
48

49
50 141 **B2:** compare U_m and U_d getting an estimation of the distance D_S between them.
51

52 142 **B3:** change SSM parameters trying to decrease D_S , go to B1 or stop and go to B4 (changing of the
53
54
55 143 parameters and stopping criteria depend on any particular realization of the optimization routine).
56

57 144 **B4:** simulate SSM with the “best-fit” parameter values corresponding to the smallest found D_S .
58

59
60 145 **B5:** generate morphological clones using the best-fit SSM with different random number sequences.
61
62
63
64
65

146

1

2 147 At the preparation stage, the QSM is formed from the TLS point cloud (A1). The detailed description

3

4 of this process is reported in [18, 19]. The resultant QSM contains all geometrical and topological

5

6 148 features needed to form the empirical distributions U_d . The distributions can be formed from several

7

8 149 tree individuals if they are close by shape to ensure the sample size of the data sets (e.g. a tree has a

9

10 150 single main stem/trunk, hence its features are underrepresented).

11

12 151

13

14 152

15

16 153 At the main cycle of the algorithm, the empirical distribution U_m is formed from the simulated SSM

17

18 154 tree (B1). Next, U_m is compared against U_d using the measure of distance (B2). The optimization

19

20 155 routine iteratively minimizes the distance value every time changing the parameter values of SSM

21

22 156 (B3), simulating SSM, and repeating the cycle from B1. After the stopping criteria of the optimization

23

24 157 routine (number of iterations, minimal allowed distance etc.) are met, the algorithm stops and produces

25

26 158 the best-fit SSM tree (B4). The best-fit SSM with different random sequences produces different

27

28 159 morphological clones.

29

30 160

31

32 161 Below we describe general aspects of each of the main components of the algorithm. The Methods

33

34 162 section addresses further the technical details.

35

36 163

37

38 164 **Quantitative Structure Model (QSM)**

39

40 165 QSM is derived from the point cloud obtained by TLS. Essentially, QSM is a surface reconstruction of

41

42 166 the branches of the real tree measured by TLS. The reconstruction itself is a stochastic process, giving

43

44 167 different architecture results for different runs. Therefore, the reconstruction introduces internal errors

45

46 168 in addition to the TLS measurement errors. Besides giving spatial locations of parts of the tree, QSM

47

48 169 also reconstructs topological relations between the tree branches. The branches of QSM consist of

49

50

51

52

53

54

55

171 elementary units, i.e. circular cylinders, but other geometrical primitives can also be applicable [31].
1
2172 Thus, any potential structural information about the original tree can be approximated with high
3
4
5173 accuracy with QSM. The details of the reconstruction algorithm are presented in [18, 19], while the
6
7174 validation of the algorithm is show in [19, 23-25]).
8

9
10175

11
12176 **Figure 2: The target QSM structure in three main 2D projections.**
13

14
15177

16
17
18178 In this work, we use the reconstructed QSM of a maple tree (Fig. 2). The QSM was selected due to its
19
20179 non-trivial form and obvious irregularities in the tree growth. This is needed to determine whether the
21
22180 stochastic rules of SSM growth can account for this variability. In fact, the QSM growth irregularities
23
24
25181 might come from some deterministic sources, like constant wind, shading from the neighbors, animal
26
27182 influences etc. Thus, our algorithm tries to compensate the lack of knowledge of the growth process
28
29
30183 with simple stochastic rules of SSM and optimization of the stochastic distance function.
31

32184

33
34
35185 **Stochastic Structure Model (SSM)**
36

37186

38
39
40187 SSM is a simulated model, preferably based on analytical and/or heuristic rules for the tree growth;
41
42188 however, any viable algorithm for generating tree forms may be used. Importantly, the ultimate output
43
44189 of the SSM simulation is a table containing data sets U , describing the tree structure.
45

46
47190

48
49191 Additionally, SSM may be supplied with stochastic variability in its parameter values. Through our
50
51
52192 studies we implement simple stochastic variations in the form of normal and uniform distributions
53
54193 added to the parameter values of SSM.
55

56
57194

58
59
60
61
62
63
64
65

195 Finally, the elementary units, such as cylinders [31], forming the SSM branches should be compatible
1
2
3 196 with the units used in the QSM reconstruction.

4
5 197
6
7 198 Examples of SSM are: *LIGNUM* [13] – a functional-structural plant model based on the physiological
8
9
10 199 principles of growth of pine trees, but also applicable to other tree forms [32]; *self-organizing tree*
11
12 200 *model* [3] based on the heuristic principles of growth, capable of producing various tree shapes and
13
14
15 201 used in computer graphics; *plastic trees* [4] procedural growth models used in computer graphics;
16
17 202 *AMAP/GreenLab* (see e.g. [33, 34]) modeling approach to generate FSPM based upon empirical rules
18
19 203 of growth with some physiological processes taken into account.
20

21
22 204
23
24 205 In this work, we use the self-organizing tree model (SOT) with shadow propagation algorithm [3] as
25
26
27 206 SSM adapted for comparison with QSM. Note that more specialized tree growth models designed for
28
29 207 the species in question would be better for the morphology optimization and the usual choice is
30
31
32 208 FSPM's, see e.g. [30].

33
34 209
35

36 210 **Structural data sets (U)**

37
38
39 211
40
41 212 Structural data sets for any given tree structure are empirical collections of the physical dimensions
42
43
44 213 and spatial orientation measures of segments and branches that are composed of segments. These data
45
46 214 sets must be similarly obtained for any pair of $\{U_m, U_d\}$ in the calculation of the structural distance.
47

48
49 215
50
51 216 Quantities in the data sets may represent scalar characteristics and/or relations between several
52
53
54 217 covariates (e.g. radii, lengths, angles, tapering function of a branch etc.). On one hand, one needs to
55
56 218 exhaustively describe morphology of the tree using various geometrical and topological features. On
57
58 219 the other hand, as the number of compared data sets $\{U_m, U_d\}$ grows the efficiency of the optimization
59

220 routine decreases, since the number of distance measures to be minimized grows correspondingly (one
1 distance value for each pair $\{U_m, U_d\}$). In case of multiple data sets, we minimize the average value of
2
3
4
5222 the corresponding distances.
6
7223
8
9
10224 Branch- and segment-related data are described in Table 1 and Fig. 8 (in Methods). These features are
11
12225 not exhaustive and can be augmented when needed. But we found this set sufficient for obtaining
13
14
15226 realistic tree shapes. Throughout the manuscript we maintain the notations B^w and S^w for the branch-
16
17227 and segment-related data sets of the (Gravelius) order w , respectively. The zero order w is assigned to
18
19228 the trunk (a branch connecting a tree with the ground). At the branching points, the lateral buds give
20
21
22229 rise to branches with order $w+1$, where w is the order of the parent branch, while the apical buds
23
24230 continue the branch of the same order.

25
26
27231

28
29232 **Table 1: Branch and segment features.**

Branch features, units	Description
β , degree	Inclination angle of the branch, i.e. angle with its parent branch.
α , 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.
Segment features, units	Description

R, m	Radius of the segment.
L, m	Distance from the beginning of the branch to the segment.
γ , degree	Angle between horizontal projections of the segment and its parent.
ζ , degree	Angle between vertical projections of the segment and its parent.

10233

11

12234 The additional details of the data set representation are explained in Methods.

13

14235

15

16236 **Measure of structural distance (D_S)**

17

18237

19

20238 The distance D_S between any two data sets, or empirical distributions, measures the difference between

21

22239 the local densities of points in U -space for these data sets, i.e. S and B tables of morphological features.

23

24240 Here, it is constructed by measuring SSM vs. QSM difference of the normalized cumulative

25

26241 distributions of the point densities projected onto a number of line directions in the coordinate space of

27

28242 the variables in U (Fig. 9A, in Methods). The difference between the projected cumulative

29

30243 distributions is further measured by the Kolmogorov-Smirnov statistic. The resulting distance between

31

32244 the two data sets U is an average of all statistics calculated from each of the lines.

33

34245

35

36246 In order to provide a reference to traditional measurement systems, we also calculate three main tree

37

38247 characteristics that are used for describing a tree shape [35], that is *height* (h), *girth* (g), and *crown*

39

40248 *spread* (c). Finally, to compare SSM and QSM shapes we calculate relative error distances d_h , d_g , and

41

42249 d_c for height, girth, and crown spread, respectively. The classical distance d_i shows how large is the

43

44250 difference between entities i of the two trees in proportion of the corresponding reference/QSM tree

45

46251 value.

47

48252

49

50253 The details of the distance calculations are further explained in the Methods section.

51

52

53

54

55

56

57

58

254

1

255 Optimization routine

3

4

5256

6

7257 The measure of structural distance $D_S(U_m, U_d)$ is minimized by adjusting the parameters ν of SSM.

8

9

10258 With infinite sampling, $D_S = 0$ for two trees that have exactly the same parameters ν . These trees are

11

12259 not copies of each other, but they are structurally similar. The choice of U defined in D_S is not unique,

13

14260 but ideally U should satisfy the following uniqueness condition for D_S to yield an acceptable measure

15

16261 of distance. Let three trees be defined by the corresponding parameters ν_A , ν_B , and ν_C with the data sets

17

18262 U_A , U_B , and U_C , respectively. Then, if $D_S(U_A, U_B) < D_S(U_A, U_C)$, one can update $C \leftarrow B$, i.e. substitute

19

20263 tree C with tree B , find any new ν_B for which the inequality holds, and repeat until $D_S(U_A, U_B) \rightarrow 0$ and

21

22264 $\nu_B \rightarrow \nu_A$. In practice, this should be true in a large neighborhood of ν_A ; however in practice, $D_S > 0$ due

23

24265 to the finite sampling and insufficient model.

25

26266

27

28267 III. Testing of the algorithm

29

30

31268

32

33269 First, we run the optimization within each of the parameter groups $I - V$, representing different

34

35270 processes of growth (see Methods for details), to determine the basic values of the parameters. These

36

37271 basic values represent choices that generate a viable tree structure similar to the target QSM. Each

38

39272 optimization run takes the best parameters for the group optimized at the previous step. The target

40

41273 structural distributions U for these runs are segment-related (S) features of the branches of topological

42

43274 order $w = 0, 1$, i.e. $S^{0,1}$. Note that this exercise serves a basic exploration of the model's behavior,

44

45275 which can be (partially) replaced, for example, by the expert guesses for the parameter values or some

46

47276 calibration process.

48

49277

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

278 Second, based on these preliminary results we may want to determine the most influential parameters
1
2279 for each of the group and combine them in a single optimization set up. Changes in these parameters
3
4
5280 cause the largest relative changes in the structural distance value. This step is required to reduce
6
7281 optimization time, and it is not needed if one possesses large enough computational resources. Several
8
9
10282 independent optimization runs were taken in order to determine the most influential parameters. For
11
12283 example, we found that the angular properties vary the least among these runs, whereas the apical
13
14
15284 dominance requires subtler adjustments (as can be understood from the complex structure of the target
16
17285 QSM).

18
19286
20

21 22287 **Low order topological adjustment of the shape**

23
24288
25
26
27289 After these initial manipulations, we obtained a model with 11 parameters and good fit of the trunk
28
29290 ($w = 0$) and first order branches (Fig. 3C) with classical metrics $d_h = 0.05$, $d_g = 0.42$, $d_c = 0.57$.
30
31
32291 However, the overall form of the resulting minimal score tree does not resemble the target QSM due to
33
34292 its rosette-shape (Fig. 3A, B). A closer look at the tree reveals that the higher order branches ($w > 1$)
35
36293 are mainly responsible for the formation of the rosette-shape of the tree, i.e. the orders which were not
37
38
39294 subject to the optimization (Fig. 3A, E). This example demonstrates the contribution of the higher
40
41295 order branches to the overall tree shape, which suggests using the information at $w > 1$ in further
42
43
44296 optimization steps. Moreover, the branch-related (B) features, such as the angular properties of
45
46297 branches of order $w > 1$, were not captured well (Fig. 3E), although similar order segment-related S -
47
48
49298 features show correct stochastic tendencies (Fig. 3D) generated automatically by the growth algorithm
50
51299 of the SSM. However, note that these features of $w > 1$ were not subject to optimization.

52
53300
54

55
56301 **Figure 3: The rosette-shape SSM resulting from adjustment of the low order segment-related**
57
58302 **scatters.** (A) The SSM tree; (B) the target QSM; (C) some segment related ($S^{0,1}$) scatters used in the
59

60
61
62
63
64
65

303 optimization; (D) higher order ($w = 2$) *S*-scatters (not used in optimization); (E) higher order ($w = 2, 3$)
 1
 2304 branch related *B*-scatters (not used in optimization). SSM/QSM scatters are shown in red/blue.
 3
 4
 5305
 6
 7306 **Low and high order topological adjustment**
 8
 9
 10307
 11
 12308 The increase in number of the structural feature tables is coupled with the increase in number of
 13
 14
 15309 distinct distance values, that is, each pair of tables (QSM vs. SSM) produces a distance score to be
 16
 17310 optimized. Although the optimization of the mean distance value for all tables hinders the
 18
 19311 improvement for each table separately, the low order and high order branches need to be fitted to the
 20
 21
 22312 corresponding branches of the target QSM. To reduce the number of distinct feature tables for the
 23
 24313 optimization we further utilize a set of merged data sets resulting in two joint segment- (*S*) and branch-
 25
 26
 27314 related (*B*) tables for all topological orders.
 28
 29315
 30
 31
 32316 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
 33
 34317 higher order branch variability (Fig. 4, $d_h = 0.08$, $d_g = 0.20$, $d_c = 0.68$). We observe a significant
 35
 36318 improvement to the tree form due to the correct account of the angular properties of the higher order
 37
 38
 39319 ($w > 1$) branches (Fig. 4E). The poor convergence of the branch linear dimensions (radii, lengths etc.)
 40
 41320 present in the branch-related tables might be due to the parameter choice of the model. Namely, the
 42
 43
 44321 small proportion of branches with similar R_f values (Fig. 4E) is the result of the fixed segment length
 45
 46322 we selected as a compromise between realism and computational complexity. The QSM minimal
 47
 48
 49323 segment length is close to zero, median is 0.06 m, whereas that of SSM is fixed to 0.2 m. We note the
 50
 51324 similar span of the curvature data points of SSM and QSM for $w = 1, 2$ (Fig. 4C and D). The $w = 2$
 52
 53
 54325 branch curvature was automatically generated by SSM as a result of the correct proportions of the $w =$
 55
 56326 1 branches, which were obtained during the optimization. Additionally, due to the lack of the
 57
 58
 59
 60
 61
 62
 63
 64
 65

327 orientation landmark in the feature data sets our best-fit SSM is fitted to the target QSM with accuracy
1
2328 of the rotation around Z-axis.

3
4
5329
6
7330 **Figure 4: Low and high order adjustment of the stochastic feature tables.** The best-fit SSM is
8
9
10331 obtained through optimization against $S^{0,1}$ and $B^{2,3,4}$ merged feature data sets. (A) The best-fit SSM
11
12332 tree, (B) the target QSM tree, (C) some projection scatters from S^1 , (D) S^2 projection scatters, (E) B^2
13
14333 and B^3 projection scatters.

16
17334
18
19335 **Clonal nature of the best-fit SSM**

20
21
22336
23
24337 Due to the highly discrete and stochastic nature of the tree growth, the structural distance hyper-
25
26
27338 surface in the space of the parameters is extremely variable (Fig. 5A). Hence, finding the global
28
29339 minima of such surface is not a trivial task. The classical smooth function optimizers are not suitable in
30
31
32340 this case, while stochastic discrete optimizers, like the genetic algorithm, seem to be more appropriate.
33
34341 Moreover, the hyper-surface itself is a stochastic entity changing every time the new sample of random
35
36342 numbers is used for a particular SSM growth realization. Therefore, any best-fit SSM is the best for a
37
38
39343 particular realization of this stochastic process and one needs to study variability of the tree shapes
40
41344 (Fig. 5B). We call these many realizations of the SSM growth *morphological tree clones*.

42
43
44345

45
46346 **Figure 5: Stochastic structure distance profiles in the parameter space.** (A) Three realizations of
47
48
49347 the distance hyper-surface projection along a dimensionless parameter λ of the SSM, controlling the
50
51348 apical dominance of a tree (the shown fragment of the projection with the step of 0.001 approximates
52
53
54349 30% of the allowed variability of the parameter during optimization, which was [0.0, 0.65]). (B)
55
56350 Structural distance values (with $U = \{S^{0,1}, B^{2,3,4}\}$) for 100 randomly generated SSM trees for each
57
58
59351 value of a discrete SSM parameter, i.e. number of growth iterations (red line connects the median

352 points of the distance distributions for each parameter value; blue line shows the same median distance
1
2353 profile but for the distance with $U = S^{0,1}$, see (C)). (C) Same as in (B), but $U = S^{0,1}$ (blue line is the
3
4
5354 median profile; red line is from (B)). The SSM is the best-fit SSM obtained in the experimentation
6
7355 reported in Fig. 4; the black arrow indicates the parameter value of the best-fit SSM found in the
8
9
10356 experimentation.

11
12357
13
14
15358 The structural distance profile depends not only on the parameters of the SSM, but the choice of the
16
17359 structural data sets. For example, in Fig. 5B and C the median distance profile is depicted given $U =$
18
19360 $\{S^{0,1}, B^{2,3,4}\}$ (red line) and $U = S^{0,1}$ (blue line). In the given parameter range the latter seems to be more
20
21
22361 flattened and lifted compared to the former. The addition of the $B^{2,3,4}$ data set might be seen as a
23
24362 perturbation to the distance profile changing the landscape properties (like minima). In our simulations
25
26
27363 we maintain the global parameter boundaries, which allows for a search within the full available space.
28
29364 However, we sequentially improve the model characteristics by perturbing the system, i.e. changing
30
31
32365 the parameters, their intervals, and the U data sets to address problematic parts of the SSM (like
33
34366 rosette-shape, Fig. 3) such that at every next optimization run the genetic algorithm is instructed to
35
36367 search around the previous best point using the initial ranges (see the genetic algorithm in Methods).

38
39368
40
41369 Given the considerations above about the nature of the structural distance hyper-surface, the further
42
43
44370 study of the morphological clones is needed. Specifically, the variability and plausibility of the clonal
45
46371 shapes need to be addressed. For example, the clones must be further selected as to produce realistic
47
48
49372 tree shapes, however, in our analysis we did not find any unrealistic tree sampled from the best-fit
50
51373 SSM. Additionally, the variability of the clones can be further calibrated, for instance, by the analysis
52
53
54374 of the natural/QSM clonal individuals.

55
56375

57
58376 **Morphological tree clones**
59
60
61
62
63
64
65

377

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

The main objective of our work is the generation of the morphological clones. In our pipeline, this occupies the last stage (see Fig. 1, B5). After the optimization is finished and the best-fit SSM is found, one can further randomize the outcome of SSM by letting the random number generator produce different sequences every time SSM is run. As a result, the different realizations of SSM should constitute the morphological clone generator yielding structural copies close to QSM and to each other and *varying* in fine detail of organization of their branches. In other words, the coarse-grain structure is repeated in each clone (and possibly grasps that of the target QSM), whereas the fine-grain structure varies.

Figure 6: Morphological clones generated from the best-fit SSM. The best-fit SSM was found using the higher topological order adjustments (Fig. 4) with number of growth iterations 30 (A), 26 (B), and 18 (C). The height, girth, crown spread, and classical metrics distributions are shown in (D) for the clones in (A), (B), and (C) (the total number of generated clones for each case is $n = 100$, only 6 are shown). The black horizontal line indicates the corresponding measure of the target QSM.

We demonstrate visualization of six clones for three distinct models in Fig. 6 (clones from other best-fit SSM's are provided at [36]). One can see the fine-grain variation in the structure in each panel of the figure, although the overall (coarse-grain) structure is preserved and presumably captures that of the target maple QSM from Fig. 2 (however, the models have higher branch densities than the QSM, due to the discretization of the space using the voxels of finite size, see [3]). The three models are: the one found during the optimization process (Fig. 6A), the one minimizing the sample median distance profile for $D_S(U = \{S^{0,1}, B^{2,3,4}\})$ shown in Fig. 5B (Fig. 6B) and one minimizing the sample median profile $D_S(U = S^{0,1})$ from Fig. 5C (Fig. 6C).

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

402 Out of 100 simulated clones for each case, we can see that the best-fit SSM obtained directly as the
403 optimization outcome (Fig. 6A) produces larger proportion of individual trees exhibiting the three
404 standard allometric measures closer to those of QSM (Fig. 6D). However, we argue that such simple
405 description of a tree, as using the allometric measures, cannot be exhaustive enough to capture both the
406 overall structure and its fine details.

407
408 The height statistics have the largest discrepancy amongst the three models in Fig. 6 but by the visual
409 inspection of the drawn clones one can see that this variability does not exert significant alterations of
410 the Z axis span and the trees seem to have even heights. Perhaps, the way we calculate the height of a
411 tree produces such large deviations in each particular case, which makes it a non-robust estimator for
412 modeling quality. That is, the overall branch density may not be affected if a single branch protrudes
413 higher than others, thus the overall outlook of a tree is not affected whereas its height is.

414
415 Note also that the model clones in Fig. 6A are the direct result of the optimization, i.e. the best-fit
416 SSM. The clones in Fig. 6B and C are obtained by manually adjusting the growth parameter to
417 minimize the median distance profiles from Fig. 5B and C, respectively. The growth iteration
418 parameter directly affects the tree height. Hence, the larger proportion of clones with heights closer to
419 the QSM height is achieved in case of the best-fit SSM (Fig. 6D). Additionally, this indicates that the
420 optimization implicitly accounts for such simple allometric measure as height.

421
422 Similarly, the girth estimation, although being captured correctly, produces large errors d_g , which
423 seems to be a result of variation in its linear dimensions (Fig. 6D). The girth dimension spans a small
424 proportion of the dimension of the whole tree: from several to tens of centimeters compared to meters
425 of the whole tree. This makes the girth specific error look gigantic (exceeding in some cases 100%)
426 and thus non-robust as well.

427

1

2428 The crown spread measure shows significant variation (Fig. 6D). We believe that this takes place due

3

4429 to the environment of the real tree the QSM was reconstructed from, which was not modeled

6

7430 appropriately in the SSM. Namely, the environmental effects (positions relative to the sun, as the tree

8

9431 grows in the Northern country, animals, winds, neighboring trees etc.) might cause systematic

10

12432 influences exerted on the shape of the QSM tree. These influences were not accounted for in the SSM,

13

14433 which was allowed to grow in any direction, limited by the uniform light conditions, existing branches

15

17434 of the same tree, and global boundaries of the available space. In addition to the environment

18

19435 influences, there are TLS measurement and QSM reconstruction errors, arising from the physical

20

22436 limitations of the instrumental technique and stochasticity of the QSM formation, respectively.

23

24437

25

26438 Finally, the true understanding of the variability of any measures of the morphological clones comes

27

29439 with the measurements of the real clones, i.e. trees similar in shape and/or genetically. Carrying out

30

31440 control experiments with QSM reconstructed from the real clonal individuals can only assess the

32

34441 variability. These real clone controlled experiments can further identify whether the obtained

35

36442 variability is large/small for the given species/clones and lead to the adjustment of the optimization

37

38443 parameters.

39

41444

42

44445 **IV. Performance of the algorithm**

45

46446

47

49447 We have performed several tests on the performance of the algorithm. The most computationally

50

51448 intensive parts of the algorithm are the SSM simulation and optimization, which both depend on a

52

54449 particular SSM implementation and the parameter search space.

53

56450

57

58

59

60

61

62

63

64

65

451 Computational time and data size scale linearly, i.e. $O(N)$, with the number of morphological features
1
2452 when extracted from a tree growth model.
3
4
5453
6
7454 Next, we used surrogate data, namely standard normal distributions, for each of the features to assess
8
9
10455 the computational complexity of the distance algorithm. The results of this numerical assay are shown
11
12456 in Fig. 7.
13

14457
15
16
17458 **Figure 7: Computational complexity of the distance algorithm.** CPU time (s) vs. number of line
18
19459 projections (A), number of structural features (B), and number of samples per structural feature (C).
20
21
22460 Specification: a single 2.9 GHz core is used; where fixed, number of features is 20, number of samples
23
24461 is 1000, number of line projections is 1000 (B) and 500 (C).
25

26462

27463 **V. Bayes Forest toolbox**

28

29463
30
31
32464
33
34465 We developed a unified interface using Matlab (MATLAB, RRID:SCR_001622) to facilitate
35
36
37466 exploration, drawing, optimization, and simulation of SSM and QSM as well as to study the
38
39467 morphological tree clones. Our interface allows for faster and easier manipulation of the required data,
40
41
42468 models, and optimization routines from the Matlab Optimization Toolbox, using only the required
43
44469 elements of otherwise complex Matlab configuration for the analysis.
45

46470
47
48
49471 The Bayes Forest toolbox is freely available at [36] (the version used in this study) and at [37] (the link
50
51
52472 is preferred for contributions and contains the latest version of the package). We also encourage the
53
54473 plant and computer scientists' community to expand their efforts using the toolbox with other species
55
56474 and models. Such a systematic approach can further be useful in tinkering the best options for creating
57
58
59475 QSM, SSM, and construction of the structural data sets.
60

476

1

2477 **VI. Discussion**

3

4

5478

6

7

8479

9

10480

11

12481

13

14

15482

16

17483

18

19484

20

21

22485

23

24

25486

26

27487

28

29488

30

31

32489

33

34490

35

36

37491

38

39492

40

41

42493

43

44494

45

46495

47

48

49496

50

51497

52

53

54498

55

56499

57

58

59500

60

61

62

63

64

65

In this work, we described an algorithmic pipeline aimed at producing stochastic structural replicas, or morphological “clones”, of trees from a QSM tree (based on TLS data) and a complimentary SSM tree. The pipeline is based on an iterative minimization of a distance between morphological structures. The distance is based on construction of the structural data sets of the tree morphologies and subsequent measure of their discrepancy using the ideas of distribution tomography analysis. The resulting best-fit morphological clones are statistically similar which is expressed in overall similarity of their form, but difference in fine details of structural organization.

Here, we have shown the general logic behind the pipeline for generation of the morphological clones. For this purpose we used a highly variable procedural tree model [3], which is more difficult to optimize. As the pipeline consists of several elementary steps, each of which can be changed according to the application and target analysis, we have proposed an initial set-up and basic configuration. We assume larger possibilities of exploration of the proposed configuration, let alone changing the steps and individual algorithms within the pipeline, which could be fulfilled by the community of plant science researchers.

The interest of building this pipeline was driven for biological applications rather than visualization purposes. Thus, for example, we use real TLS measurements and general-purpose measure of the distance, while omitting visual effects (e.g. shades, leaves etc.). We believe this pipeline can be useful in the rigorous analysis of the plant morphogenesis and corresponding applications, which differ from similar studies done in computer graphics field, e.g. [5].

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

501 Moreover, our algorithm makes use of the distance measure taking into account significant portion of
502 the data for at least one topological order. This allows for a more comprehensive analysis of forms and
503 their description, using empirical distributions of morphological features rather than scalar allometric
504 entities. Due to this reason, we do not rely on the traditional metrics comparison in this work as we
505 found that similar values for the height, girth, and crown distances may correspond to different tree
506 forms and, thus, be non-robust.

507
508 Use of several QSM trees can enhance the robustness of the statistical analysis presented here. In this
509 case, similarly looking trees should be used and the degree of similarity might be established using our
510 definition of the structural distance. For example, the trunk features are more reliably reproduced in
511 statistical sense, when several QSM's are used. It might be stressed that other notions of "clone" can
512 be used to establish relationship with morphology. Thus, the genetic clones might be utilized to
513 establish to what degree the morphology of a tree is encoded into genes.

515 **Methods**

517 **Laser scanning measurements**

518
519 The subject tree was measured in leaf-off conditions and our system consisted of a phase-shift based
520 terrestrial laser scanner, namely the Leica HDS6100 with a 650–690 nm wavelength. The distance
521 measurement accuracy and the point separation angle of the scanner were about 2–3 mm and 0.036
522 degrees, respectively. The horizontal distance of the scanner to the trunk was about 7–12 m, giving an
523 average point density on the surface of the trunk (at the level of the scanner) for a single scan around
524 2–5 points per square centimeter. The QSM of the subject maple tree consists of 19,000 cylinders
525 approximating 3,078 branches.

526

1

2527 **Self-organizing tree model (SOT)**

3

4

5528

6

7529 In this work we used SOT implemented in the LPFG simulator, part of the Virtual Laboratory software

8

9530 suite [38], version 4.4.0-2424 for 64-bit Mac OS, see [39]. This procedural tree model is fast and able

10

11531 to generate variety of forms.

12

13532

14

15533 The total number of growth parameters of the model is 27: 23 are grouped, 4 are fixed. The values of

16

17534 the latter are dictated both by suggestions of the authors in [3] and the compromise between

18

19535 computation time and details of the morphological description. For example, the segment length is

20

21536 0.2 m, the voxel size is 0.2 m, and the model tree grows within 12x12x12 m cube from the center of

22

23537 XY bottom plane of the cube (Z-axis is oriented upwards).

24

25538

26

27539 The grouped parameters are divided between 5 distinct groups corresponding to different related

28

29540 processes:

30

31541 *Group I*: the initial growth parameters, including limiting values, and pipe model related parameters.

32

33542 *Group II*: environmental effects such as sensitivity to the neighborhood shading, vertical gradient

34

35543 distribution of the light, tropism etc.

36

37544 *Group III*: apical dominance parameters.

38

39545 *Group IV*: shadow propagation related constants (see [3]).

40

41546 *Group V*: angular/branching properties.

42

43547

44

45548 **Data set representation**

46

47549

48

49

50

51

52

53

54

55

56

57

58

550 One needs a more compact representation of the data, since the larger number of data sets U is, the
1
2551 larger number of distance values is and the more difficult the optimization becomes. One solution is to
3
4
5552 use larger data sets with all application specific features. Therefore, we use tables including all
6
7553 measured features; hence, one table represents a data set. However, it is not possible to merge
8
9
10554 segment- and branch-related features into a single table as these differ in dimension (usually one
11
12555 branch is composed of many segments). Thus, we usually compare the array of pairs $\{U_m, U_d\}$, having
13
14
15556 as a result the array of distance values, but with such larger table representation we have smaller size
16
17557 of these arrays. For example, one could form U_1 from inclination angle β and U_2 from azimuthal angle
18
19
20558 α , or, alternatively, from a larger table U consisting of all branching features (β , α , L_t , R_f , L_a from
21
22559 Table 1). The former distance would be formed from D_1 and D_2 for U_1 and U_2 respectively, whereas
23
24
25560 the latter case would require a single distance to optimize.

26
27561
28
29
30562 **Figure 8: Visual structure of a tree and its representation using the structural data sets U .** (A) A
31
32563 sample tree; (B) geometrical features of the branch- (B) and segment-related (S) data sets; and (C)
33
34564 various projections of the U data sets.

35
36
37565
38
39
40566 Additionally, it is possible to merge the corresponding data sets for several topological orders, which
41
42567 results at most in two large data sets of branching and segment features, respectively. While this
43
44
45568 simplifies the search of the distance minimum, this technique must be used with care as in this case
46
47569 one heavily relies upon the growth rules of SSM. If these rules are not based on biologically motivated
48
49
50570 rules, SSM can produce highly unrealistic tree forms as the “best-fit”, since there is a possibility to mix
51
52571 the features of different topological orders. For example, the branches of higher order could be much
53
54572 thicker than those of the lower order, which should not happen using biologically based growth
55
56
57573 algorithms (e.g. pipe model).

58
59574
60
61
62
63
64
65

575 In a simulated SSM structure the extraction of topological relations between branches is
1
2576 straightforward: the lateral buds start the next order and apical buds continue the current order.
3
4
5577 However, this is not the case with QSM since it is a time snapshot of a tree form that does not retain
6
7578 the history of the tree growth. Thus, the reconstruction algorithm requires other means for extracting
8
9
10579 the topology. Although the reconstruction algorithm defines a complicated procedure that outlines the
11
12580 topology of a tree, it can be roughly approximated by the following rule: at branching points the
13
14
15581 thickest branch is the continuation of the same order w , while thinner branches are lateral expansions
16
17582 of the order $w + 1$ [18]. For the species with weak apical dominance (shrubby trees) we follow the rule
18
19583 when extracting topology from SSM (for the species with strong apical dominance, the rule gives the
20
21
22584 same result as the analytical procedure).

23
24585

25

26

27586 **Structural (D_S) and traditional distances**

28

29587

30

31

32588 The structural distance is calculated measuring difference between the normalized cumulative

33

34589 distributions of two point densities projected onto a number of line directions. The directions of lines

35

36590 are generated with quasi-Monte Carlo method using low-discrepancy (quasi-random) sequences, which

37

38

39591 cover the given space more evenly than uniformly generated sequences.

40

41592

42

43

44593 The empirical probability density function $p(U)$, $U \in R^N$, can be approximated by the series of 1D

45

46594 density functions $p_{1D}(U, L)$, where L is a line in R^N . Each of these 1D functions is constructed by

47

48

49595 projecting all the data points of U onto a line L (in this work we used 1000 lines). Cumulative

50

51596 distributions $P_{1D}(U_m, L_i)$ and $P_{1D}(U_d, L_i)$ (U_m and U_d being the two point densities, e.g. SSM vs. QSM

52

53

54597 data sets) for each line direction L_i are compared, and for any given data set pair $\{U_m, U_d\}$ the resultant

55

56598 distance value is:

57

58

59

60

61

62

63

64

65

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

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

Figure 9: Distribution tomography of the structural data sets (A) and classical metric for the crown spread (B). (A) Data points in U (projected here for simplicity onto (u_i, u_j) plane, i.e. in 2D) are used to construct the projection onto a line L . Cumulative empirical distribution is calculated along L (red). Only one line is shown. (B) Top view of a tree: spokes (red) emanate from the ground segment (green) extending up to the most distant points (blue).

Height (h) is calculated as the highest point of a tree. *Girth* (g) is calculated as the diameter of the ground segment, because the diameter at breast height is not appropriate for the shrubby trees. *Crown spread* (c) is calculated as follows. First, on XY-plane (top view, Fig. 9B) the set of spokes (red lines in Fig. 9B) emanating from the center of a tree (the ground segment, green circle) is built with 10 degrees azimuthal separation. Then the length of each spoke is calculated as a distance from the tree center to the most distant point of the crown in the direction of the spoke (blue circles). The crown spread is twice the average of all spokes of a tree.

Finally, when comparing two tree shapes with traditional metrics we calculate the distances as follows:

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

In this, h_d , g_d , and c_d are the height, girth, and crown spread of the QSM tree, respectively, whereas h_m , g_m , and c_m are the corresponding attributes of the best-fit SSM tree.

622 Genetic algorithm

1
2623
3
4
5624 Any algorithm from a standard optimization library (e.g. Matlab Optimization Toolbox) that finds a
6
7625 minimum of an objective function ($D_S = F(v)$) can be used. However, to facilitate global minimum
8
9
10626 search and given the nature of the problem we use the genetic algorithm (implemented in Matlab,
11
12627 version R2015b). Additionally, some parameters of SSM may take only integer values, so the genetic
13
14628 algorithm handles the integer parameters correctly unlike, for example, the classical steepest decent
15
16
17629 algorithm. The genetic algorithm iteratively finds a minimum of D_S , each iteration being called
18
19630 *generation*. Each generation is characterized with a number of individuals, i.e. *population*; one
20
21
22631 individual is equivalent to one set of the parameter values. The variation is controlled by the *crossover*
23
24632 *rate* (rate of recombination of the population parameters) and *mutation rate* (rate of introduction of the
25
26
27633 new variability into the population). The former is fixed to 80% in the Matlab Optimization Toolbox,
28
29634 whereas the latter is controlled by our configuration (19% in the rosette-shape example, Fig. 3, and
30
31
32635 15% in the best-fit SSM from Fig. 4). The user controls values' ranges of the parameters. There are
33
34636 two types of ranges: *global* lower and upper boundaries for each of the parameter values and *initial*
35
36637 *range*, from which the algorithm tries to construct the initial population. The latter impacts the
37
38
39638 convergence rate: if it is too broad poor convergence is attained. Finally, the algorithm stops when it
40
41639 reaches a fixed number of generations without improving the distance.

42
43
44640
45
46641 Thus, the objective function takes the input parameters v , simulates SSM with v , calculates and returns
47
48
49642 structural data sets U_m . Subsequently, the objective function calculates $D_S(U_m, U_d)$ and returns it to the
50
51643 optimization routine. The SSM, being a stochastic model, *must* have a fixed random generator seed
52
53
54644 during optimization, i.e. the same input parameter set must produce the same structural output. This is
55
56645 needed for convergence of the optimization. After obtaining the final best-fit form of SSM, one can

646 further explore the variability coming from different random number sequences used in the SSM
1
2 647 simulations. Such random best-fit SSM is capable of producing the clonal morphologies.
3

4
5 648

6

7 649 **Availability of supporting source code and requirements**

8
9

10 650 Project name: BayesForest

11

12 651 Project home page: <https://github.com/inuritdino/BayesForest/wiki>

13
14

15 652 Operating system: Platform independent

16

17 653 Programming language: Matlab

18
19

20 654 Other requirements: VLAB software suite, version $\geq 4.4.0-2424$

21

22 655 License: MIT

23
24

25 656

26

27 657 **Data availability**

28
29

30 658 All data needed to reproduce the results of this study, some additional materials, and Bayes Forest

31
32

33 659 Toolbox are available at [36]. The most recent version of the Toolbox is also available at [37] (this

34

35 660 interface is preferred for the contributors and also contains the most recent version of the software).

36
37

38 661 Tutorials on how to run the tests using the Matlab toolbox are available in the BayesForest Wiki [40]

39

40 662 and snapshots are also available in the GigaScience repository, GigaDB [41].

41
42

43 663

44

45 664 **List of Abbreviations**

46

47 665 FSPM – functional-structural plant model.

48
49

50 666 QSM – quantitative structure model.

51

52 667 SSM – stochastic structure model.

53
54

55 668 SOT – self-organizing tree model.

56

57 669 TLS – terrestrial laser scanning.

58
59

60 670

61

62

63

64

65

671 **Ethics approval and consent to participate**
1
2672 Not applicable
3
4
5673
6
7674 **Consent for publication**
8
9
10675 Not applicable
11
12676
13
14677 **Competing interests**
15
16
17678 The authors declare that they have no competing interests
18
19679
20
21
22680 **Funding**
23
24681 This work was supported by the Academy of Finland: Suomen Akatemia (Center of Excellence in
25
26
27682 Inverse Problem Research, one of the PI's is Mikko Kaasalainen).
28
29683
30
31
32684 **Author contributions**
33
34685 IP performed all simulations, processed the data, and wrote the manuscript; MJ wrote the code for
35
36686 calculating the structural distance, discussed the results; MÅ contributed to Bayes Forest Toolbox; PR
37
38
39687 generated and provided for the QSM data, wrote the manuscript and discussed the results; MK
40
41688 conceived the study, discussed the results, and wrote the manuscript.
42
43
44689
45
46690 **Acknowledgments**
47
48
49691 We would like to thank Risto Sievänen and Wojtek Palubicki for useful discussion and comments on
50
51692 the model design and implementation.
52
53
54693
55
56694 **References**
57
58
59695
60
61
62
63
64
65

- 696 [1] Prusinkiewicz P. Modeling plant growth and development. *Current Opinion in Plant Biology*.
1
2697 2004;7:79-83.
3
- 4
5698 [2] Fourcaud T, Zhang X, Stokes A, Lambers H, Körner C. *Plant Growth Modelling and Applications:*
6
7699 *The Increasing Importance of Plant Architecture in Growth Models.* *Ann Bot.* 2008;101:1053-1063.
8
- 9
10700 [3] Palubicki W, Horel K, Longay S, Runions A, Lane B, Mech R, et al. Self-organizing tree models
11
12701 for image synthesis. *ACM Transactions on Graphics.* 2009;28:58.
13
- 14
15702 [4] Pirk S, Stava O, Kratt J, Abdul Massih Said M, Neubert B, Mech R, et al. *Plastic Trees: Interactive*
16
17703 *Self-Adapting Botanical Tree Models.* *ACM Transactions on Graphics.* 2012;31:50.
18
- 19
20704 [5] Stava O, Pirk S, Kratt J, Chen B, Mech R, Deussen O, et al. *Inverse Procedural Modelling of Trees.*
21
22705 *Computer Graphics Forum.* 2014;33:118-131.
23
- 24
25706 [6] Hallé F, Oldeman R, Tomlinson P. *Tropical trees and forests: An architectural analysis.* Berlin:
26
27707 Springer; 1978.
28
- 29
30708 [7] Sachs T, Novoplansky A. *Tree form: architectural models do not suffice.* *Israel J Plant Sci.*
31
32709 1995;43:203–212.
33
- 34
35710 [8] Room P, Hanan J, Prusinkiewicz P. *Virtual plants: new perspectives for ecologists, pathologists*
36
37711 *and agricultural scientists.* *Trends Plant Sci.* 1996;1:33-38.
38
- 39
40712 [9] Sievänen R, Nikinmaa E, Nygren P, Ozier-Lafontaine H, Perttunen J, Hakula H. *Components of*
41
42713 *functional–structural tree models.* *Ann Sci.* 2000;57:399-412.
43
- 44
45714 [10] Godin C, Hanan J, Kurth W, Lacoite A, Takenaka A, Prusinkiewicz P, et al., editors.
46
47715 *Proceedings of the 4th International Workshop on Functional–Structural Plant Models, June 7-11,*
48
49716 *Montpellier, France; 2004.*
50
- 51
52717 [11] Mäkelä A, Hari P. *Stand growth model based on carbon uptake and allocation in individual trees.*
53
54718 *Ecol Model.* 1986;33:204-229.
55
- 56
57719 [12] Rauscher H, Isebrands J, Host G, Dickson R, Dickmann D, Crow T, et al. *ECOPHYS: An*
58
59720 *ecophysiological growth process model for juvenile poplar.* *Tree Physiol.* 1990;7:255-281.
60
61
62
63
64
65

- 721 [13] Perttunen J, Sievänen R, Nikinmaa E, Salminen H, Saarenmaa H, Väkevä J. LIGNUM: a tree
1
2722 model based on simple structural units. *Ann Bot.* 1996;77:87-98.
3
4
5723 [14] Lacoite A. Carbon allocation among tree organs: a review of basic processes and representation
6
7724 in functional–structural tree models. *Ann For Sci.* 2000;57:521-533.
8
9
10725 [15] Rosell J, Llorens J, Sanz R, Arnó J, Ribes-Dasi M, Masip J, et al. Obtaining the three-dimensional
11
12726 structure of tree orchards from remote 2D terrestrial LIDAR scanning. *Agric and For Meteor.*
13
14727 2009;149:1505-1515.
15
16
17728 [16] Van Leeuwen M, Nieuwenhuis M. Retrieval of forest structural parameters using lidar remote
18
19729 sensing. *Eur J For Res.* 2010;129:749–770.
20
21
22730 [17] Rutzinger M, Pratihast A, Oude Elberink S, Vosselman G. Detection and modelling of 3D trees
23
24731 from mobile laser scanning data. In: *International Archives of Photogrammetry, Remote Sensing and*
25
26732 *Spatial Information Sciences.* 2010;XXXVIII:520-525.
27
28
29733 [18] Raumonon P, Kaasalainen M, Åkerblom M, Kaasalainen S, Kaartinen H, Vastaranta M, et al. Fast
30
31734 Automatic Precision Tree Models from Terrestrial Laser Scanner Data. *Remote Sensing.* 2013;5:491-
32
33735 520.
34
35
36736 [19] Calders K, Newnham G, Burt A, Murphy S, Raumonon P, Herold M, et al. Nondestructive
37
38737 estimates of above-ground biomass using terrestrial laser scanning. *Methods in Ecol Evol.* 2015;6:198-
39
40738 208.
41
42
43739 [20] Xu H, Gossett N, Chen B. Knowledge and Heuristic Based Modeling of Laser-Scanned Trees.
44
45740 *ACM Transactions on Graphics.* 2007;26:19.
46
47
48741 [21] Livny Y, Yan F, Olson M, Chen B, Zhang H, El-Sana J. Automatic Reconstruction of Tree
49
50742 Skeletal Structures from Point Clouds. *ACM Transactions on Graphics.* 2010;29:151.
51
52
53743 [22] Preuksakarn C, Boudon F, Ferraro P, Durand JB, Nikinmaa E, Godin C. Reconstructing Plant
54
55744 Architecture from 3D Laser scanner data. In: DeJong T., Da Silva D, editors. *Proceedings of the 6th*
56
57745 *International Workshop on Functional-Structural Plant Models.* 2010. 14-16.
58
59
60
61
62
63
64
65

- 746 [23] Hackenberg J, Spiecker H, Calders K, Disney M, Raunonen P. SimpleTree - an efficient open
1
2747 source tool to build tree models from TLS clouds. *Forests*. 2015;6:4245-4294.
3
4
5748 [24] Kaasalainen S, Krooks A, Liski J, Raunonen P, Kaartinen H, Kaasalainen M, et al. Change
6
7749 Detection of Tree Biomass with Terrestrial Laser Scanning and Quantitative Structure Modeling.
8
9750 *Remote Sensing*. 2014;6:3906-3922.
10
11
12751 [25] Raunonen P, Casella E, Calders K, Murphy S, Åkerblom M, Kaasalainen M. Massive-scale Tree
13
14752 Modelling from TLS Data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial*
15
16753 *Information Sciences*. 2015;II-3/W4:189-196.
17
18
19754 [26] Smith A, Astrup R, Raunonen P, Liski J, Krooks A, Kaasalainen S, et al. Tree Root system
20
21
22755 characterization and volume estimation by terrestrial laser scanning. *Forests*. 2014;5:3274-3294.
23
24756 [27] Kaasalainen M. Dynamical Tomography of Gravitationally Bound Systems. *Inverse Problems and*
25
26757 *Imaging*. 2008;2:527–546.
27
28
29758 [28] Bracewell R. Numerical Transforms. *Science*. 1990;248:697-704.
30
31
32759 [29] Sievänen R, Perttunen J, Nikinmaa E, Kaitaniemi P. Toward extension of a single tree functional-
33
34760 structural model of Scots pine to stand level: effect of the canopy of randomly distributed, identical
35
36761 trees on development of tree structure. *Functional Plant Biology*. 2008;;964–975.
37
38
39762 [30] Potapov I, Järvenpää M, Åkerblom M, Raunonen P, Kaasalainen M. Data-based stochastic
40
41763 modeling of tree growth and structure formation. *Silva Fennica*. 2016;50:1413.
42
43
44764 [31] Åkerblom M, Raunonen P, Kaasalainen M, Casella E. Analysis of Geometric Primitives in
45
46765 Quantitative Structure Models of Tree Stems. *Remote Sensing*. 2015;7:4581-4603.
47
48
49766 [32] Lu M, Nygren P, Perttunen J, Pallardy S, Larsen D. Application of the functional-structural tree
50
51767 model LIGNUM to growth simulation of short-rotation eastern cottonwood. *Silva Fennica*.
52
53768 2011;45:431–474.
54
55
56769 [33] De Reffye P, Fourcaud T, Blaise F, Barthelemy D, Houllier F. A functional model of tree growth
57
58770 and tree architecture. *Silva Fennica*. 1997;31:297-311.
59
60
61
62
63
64
65

771 [34] Yan H, Kang M, de Reffye P, Dingkuhn M. A Dynamic, Architectural Plant Model Simulating
1
2772 Resource-dependent Growth. *Annals of Botany*. 2004;93:591-602.
3
4
5773 [35] Frank E. A Numerical Method of Plotting Tree Shapes. *Bull East Nat Tree Soc*. 2010;6:2-8.
6
7774 [36] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M. Bayes Forest Toolbox.
8
9775 <http://math.tut.fi/inversegroup/app/bayesforest/v1/>. Accessed 25 June 2017.
10
11
12776 [37] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M. Bayes Forest Toolbox: the
13
14777 developmental version. <https://github.com/inuritdino/BayesForest>.
15
16
17778 [38] Federl P, Prusinkiewicz P. Virtual Laboratory: an interactive software environment for computer
18
19779 graphics. In *Proceedings of Computer Graphics International*. 1999, p. 93-100.
20
21
22780 [39] Prusinkiewicz P. Virtual Laboratory (VLAB) / L-studio homepage.
23
24781 http://algorithmicbotany.org/virtual_laboratory/. Accessed 17 Apr 2017.
25
26
27782 [40] The BayesForest Wiki. <https://github.com/inuritdino/BayesForest/wiki>. Accessed 24 June 2017.
28
29783 [41] Potapov I, Järvenpää M, Åkerblom M, Raumonen P, Kaasalainen M: Supporting data for "Bayes
30784 Forest: a data-intensive generator of morphological tree clones" GigaScience Database. 2017.
31785 <http://dx.doi.org/10.5524/100337>
32786
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

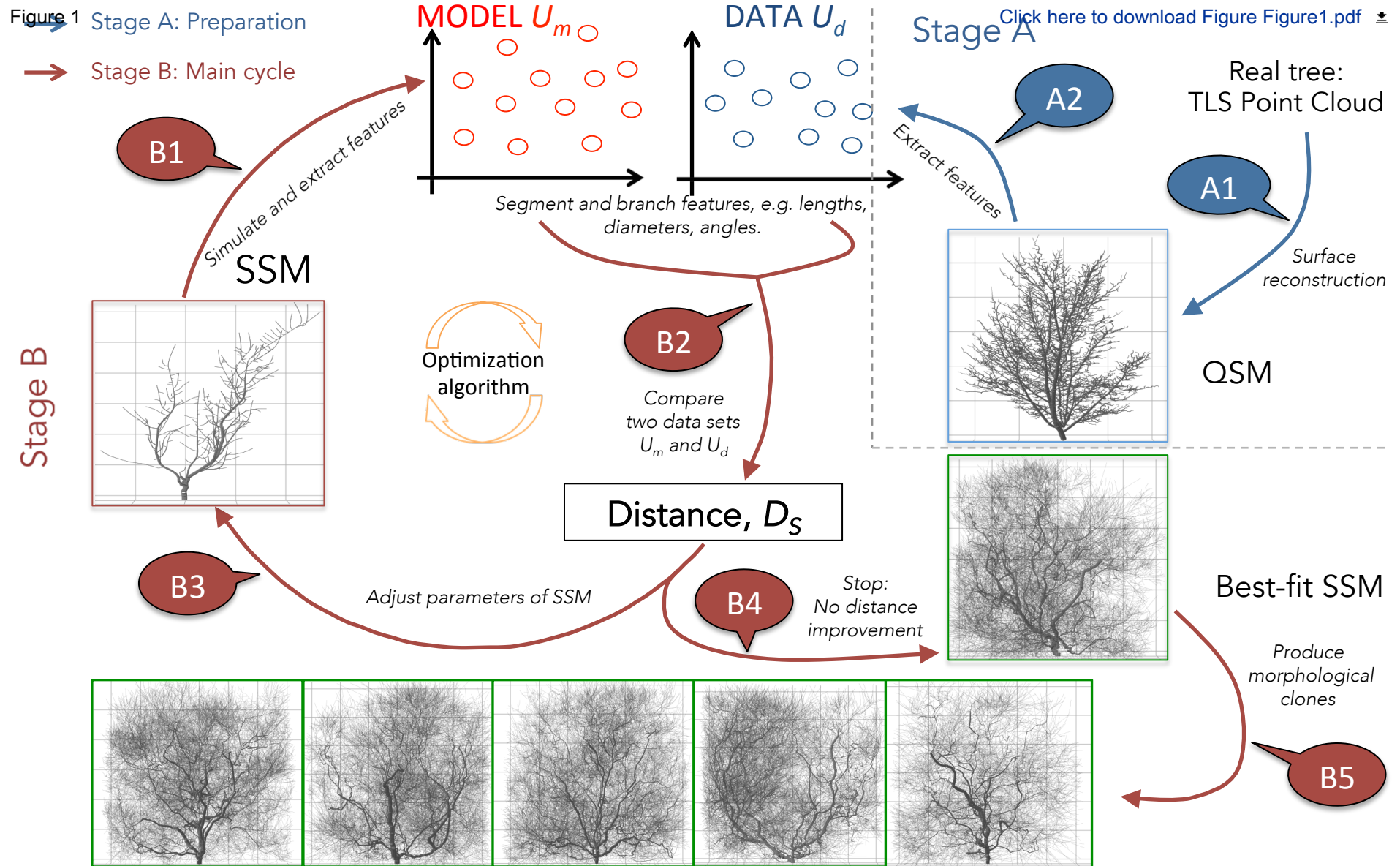
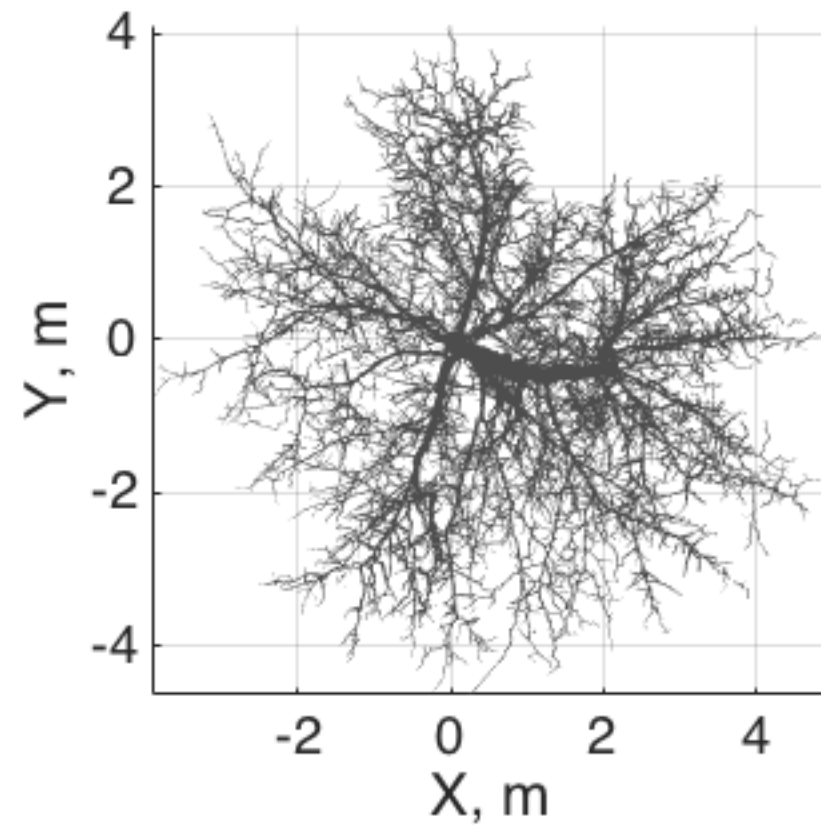
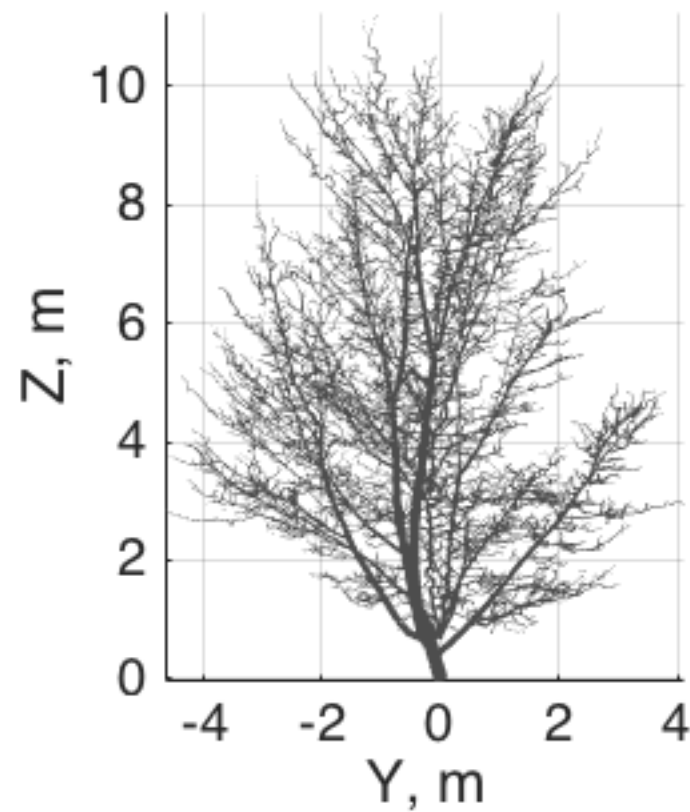
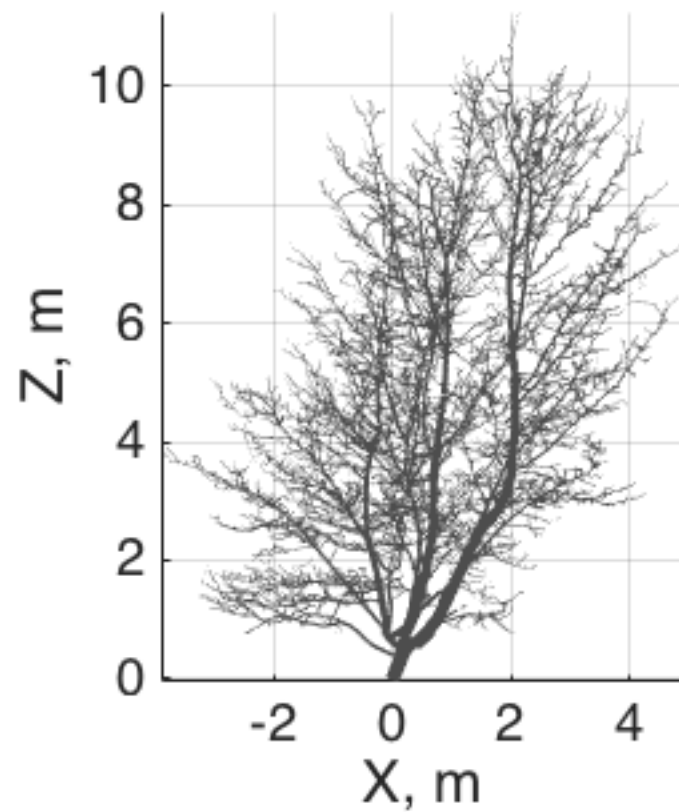
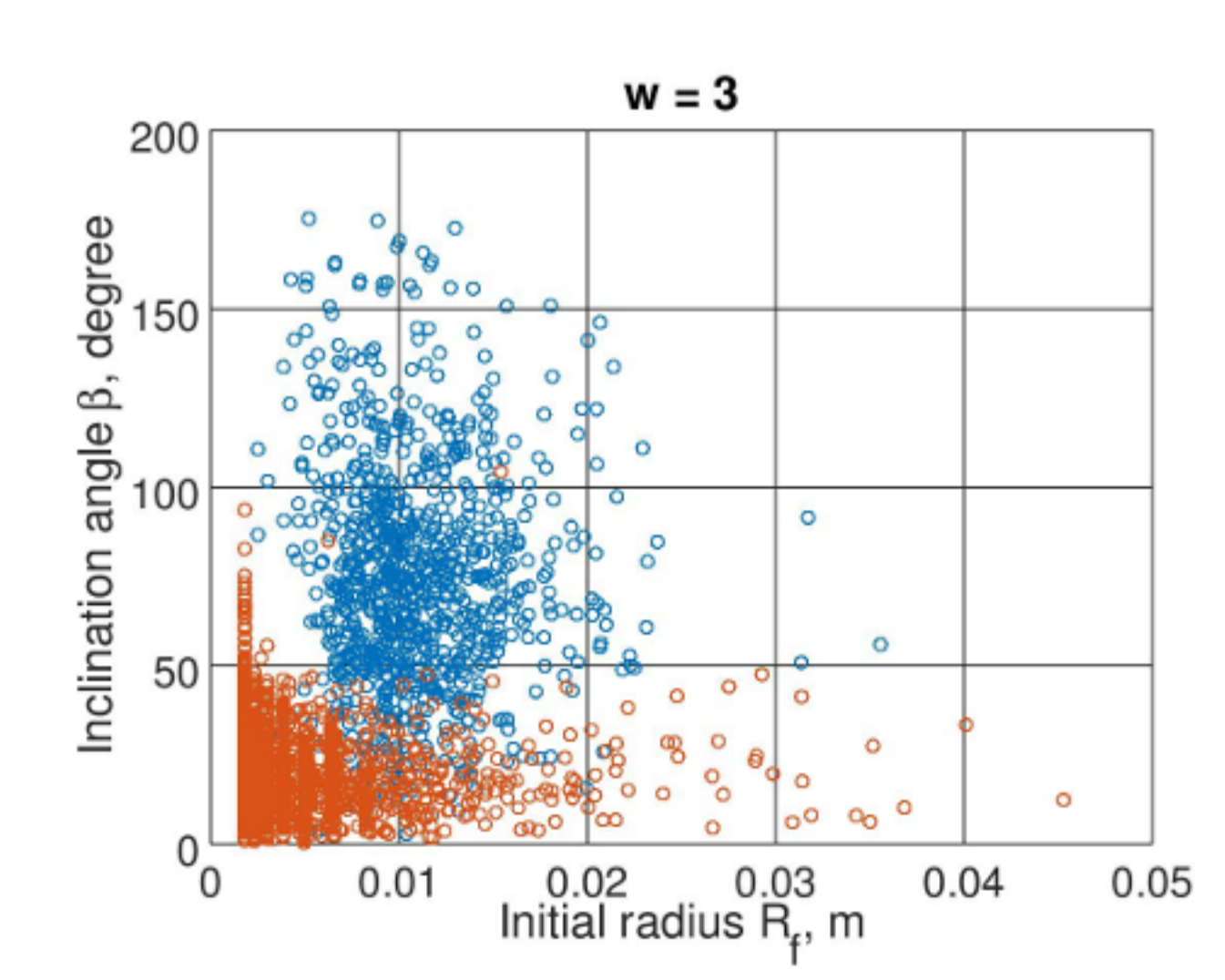
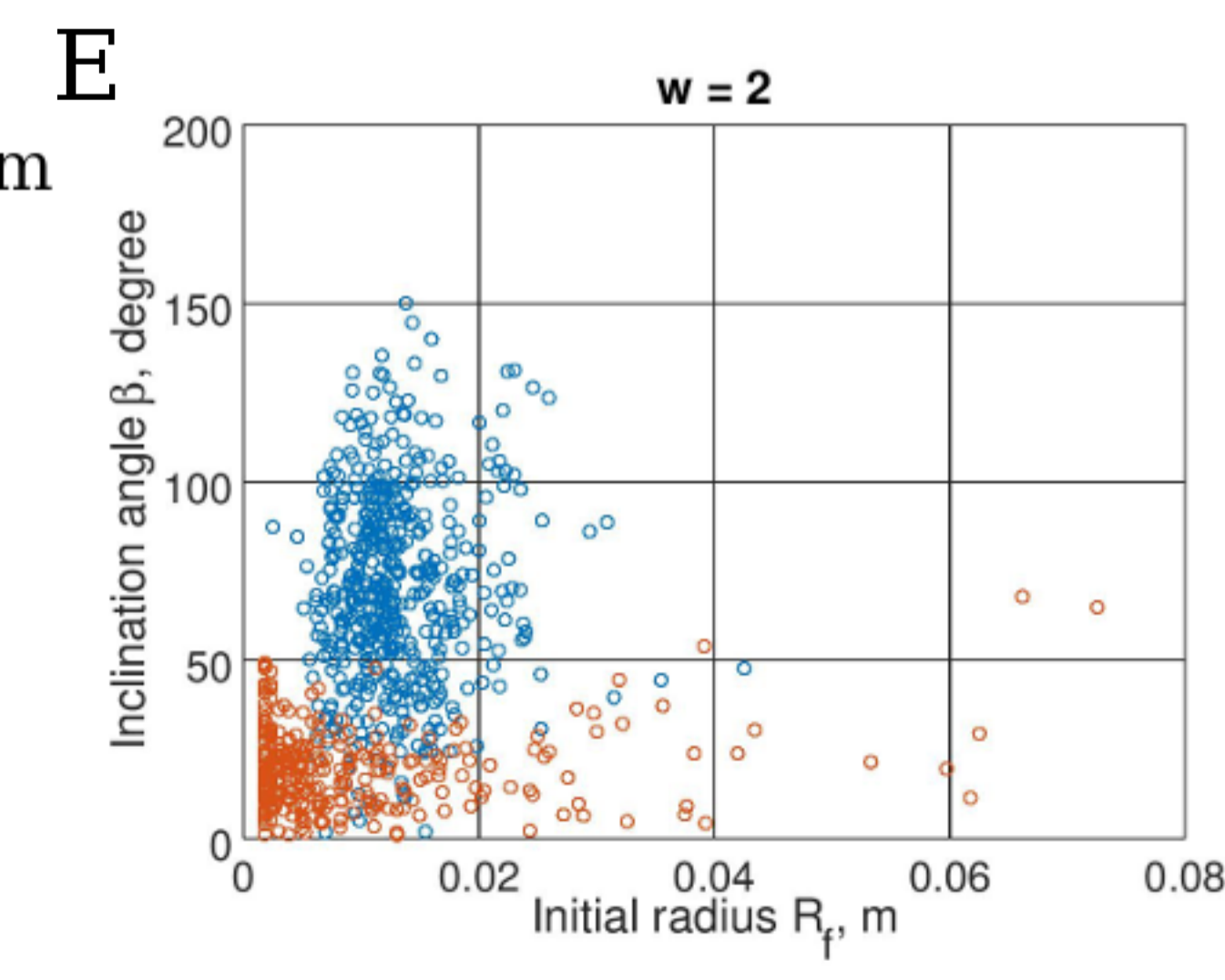
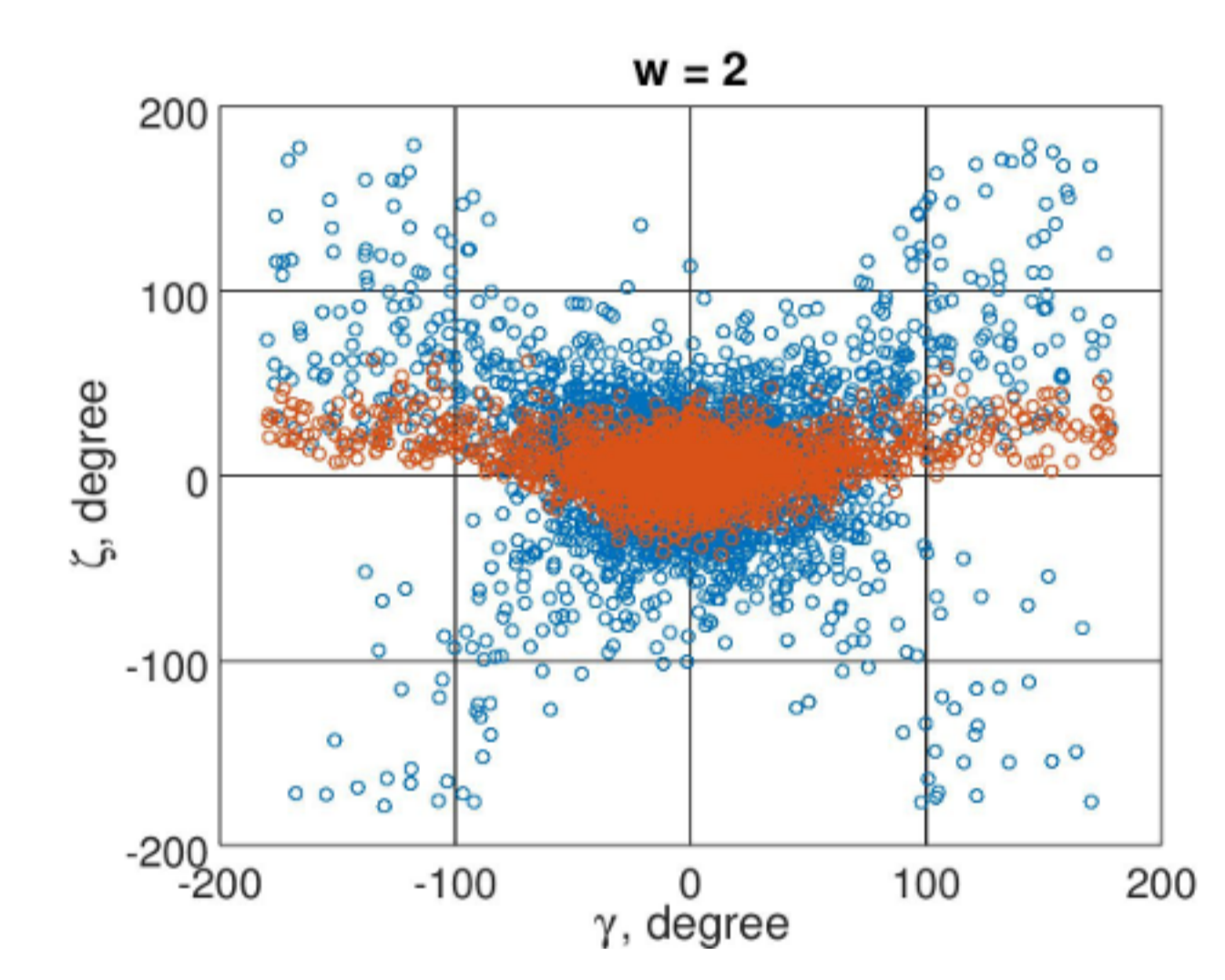
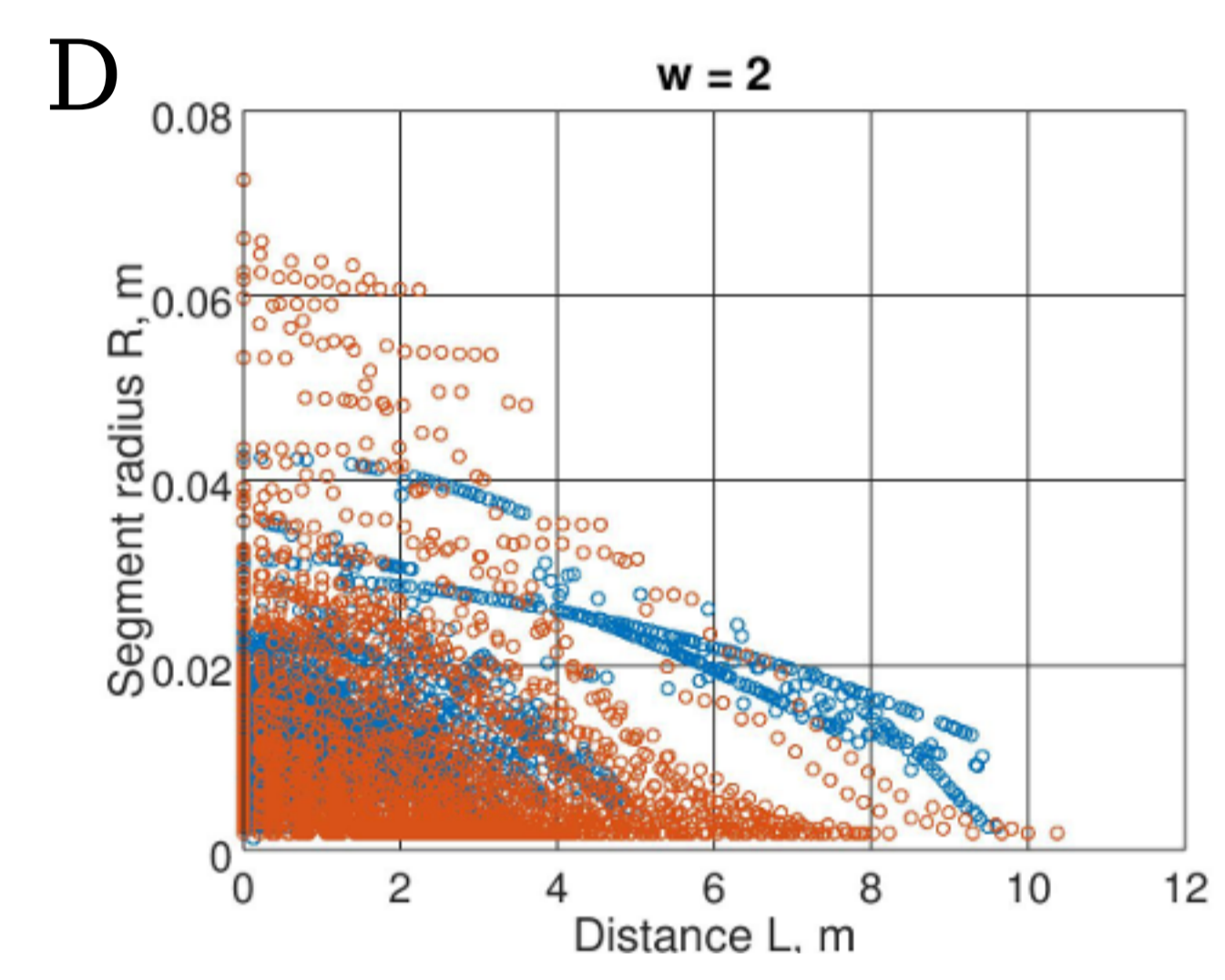
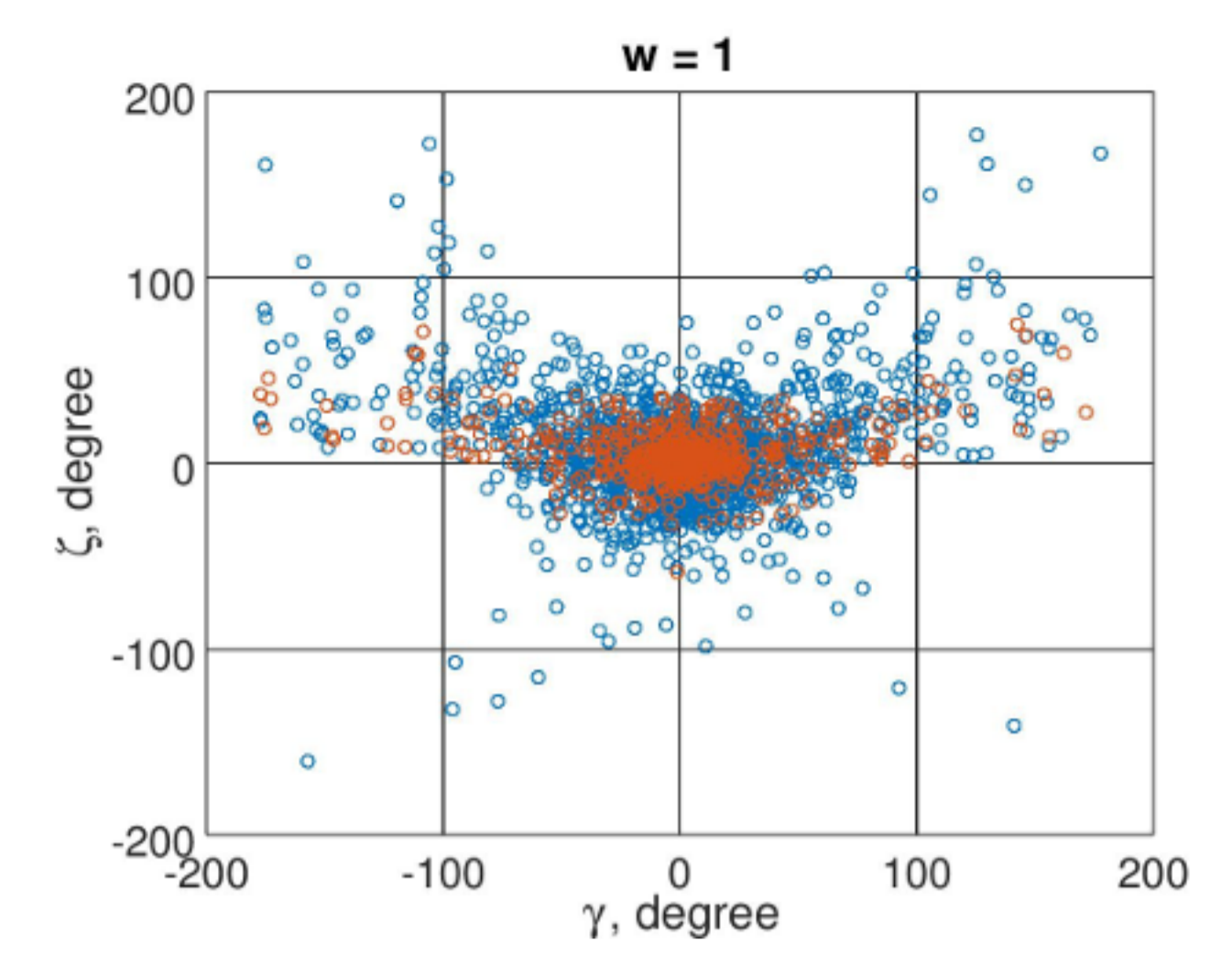
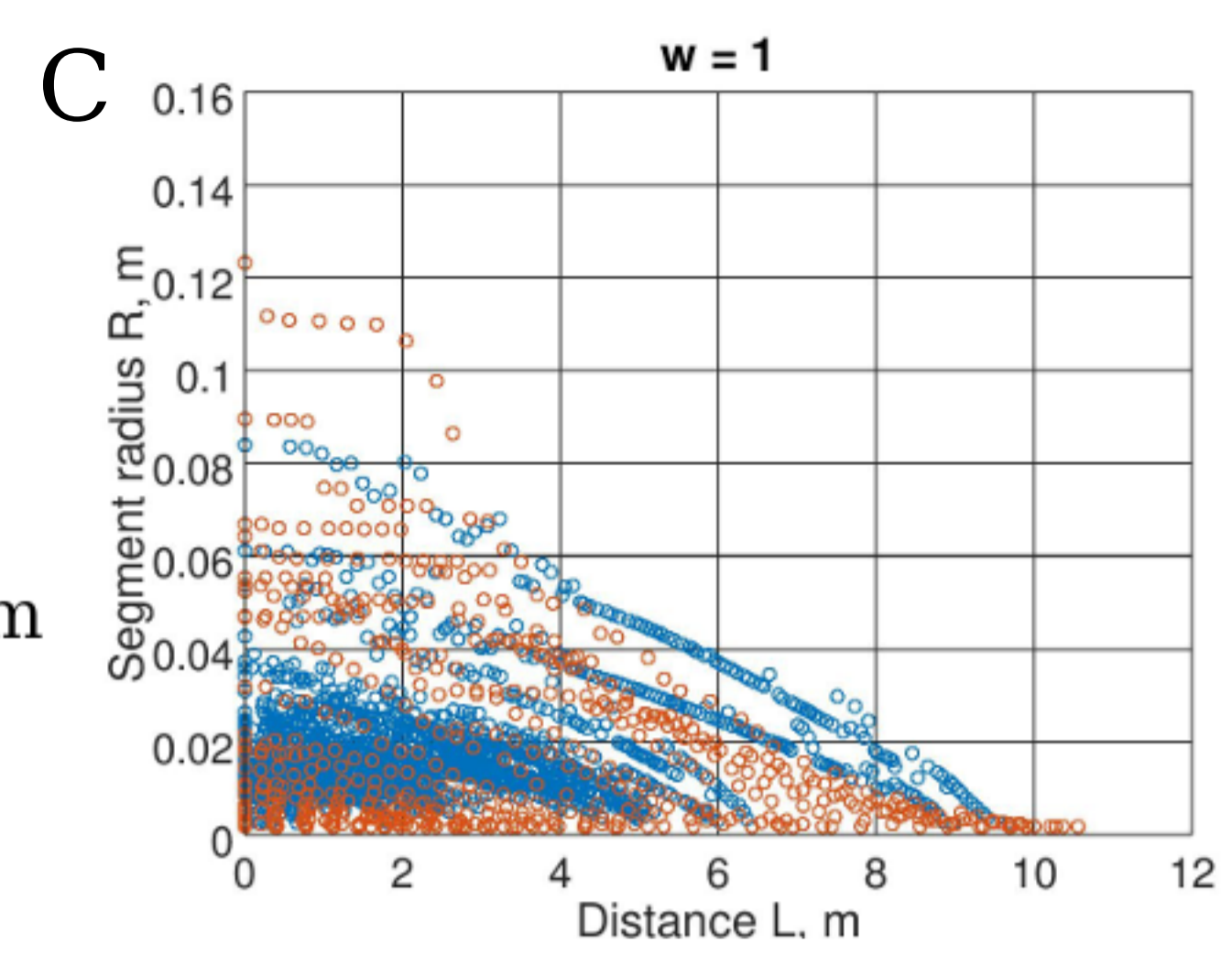
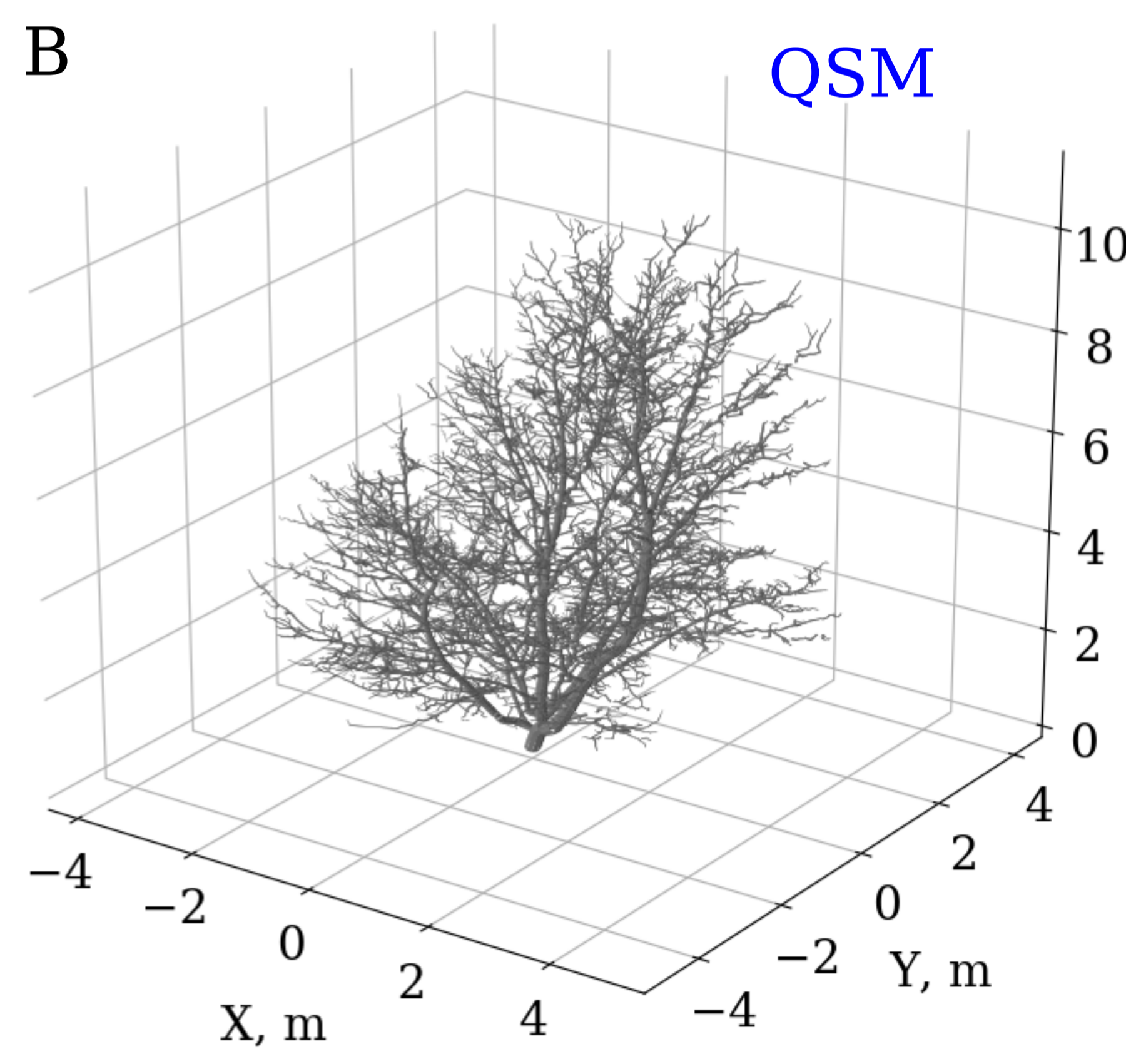
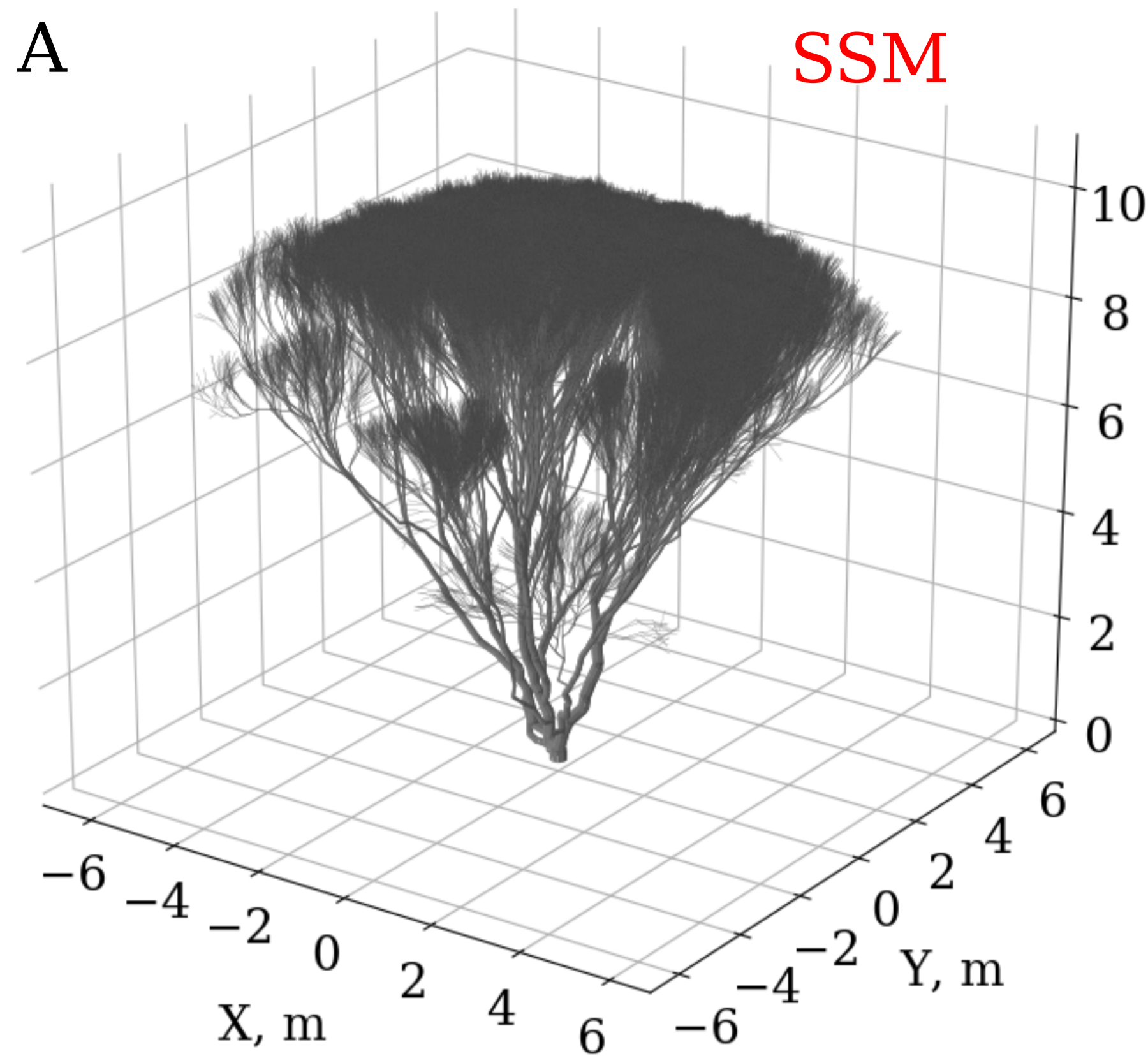
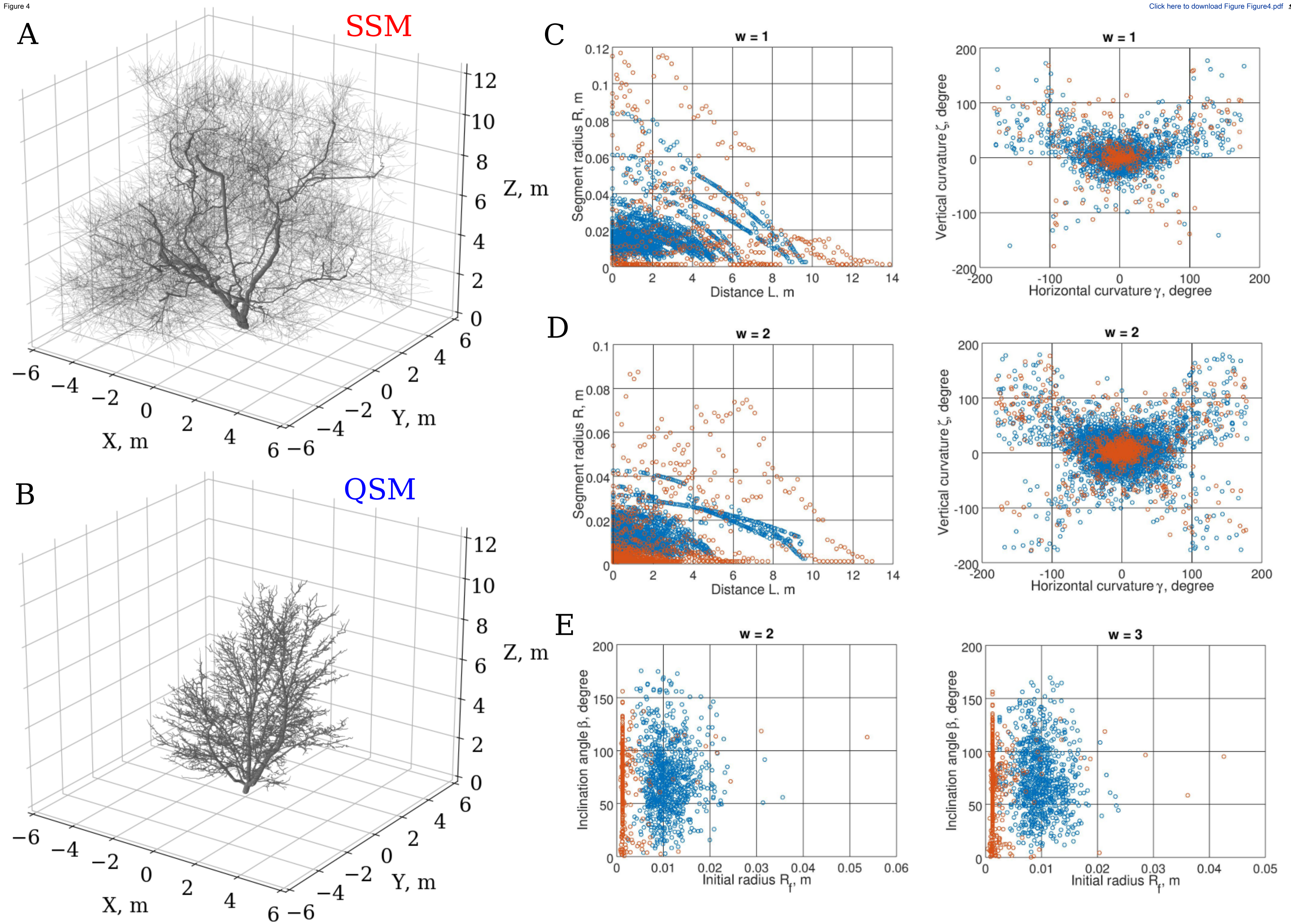


Figure 2

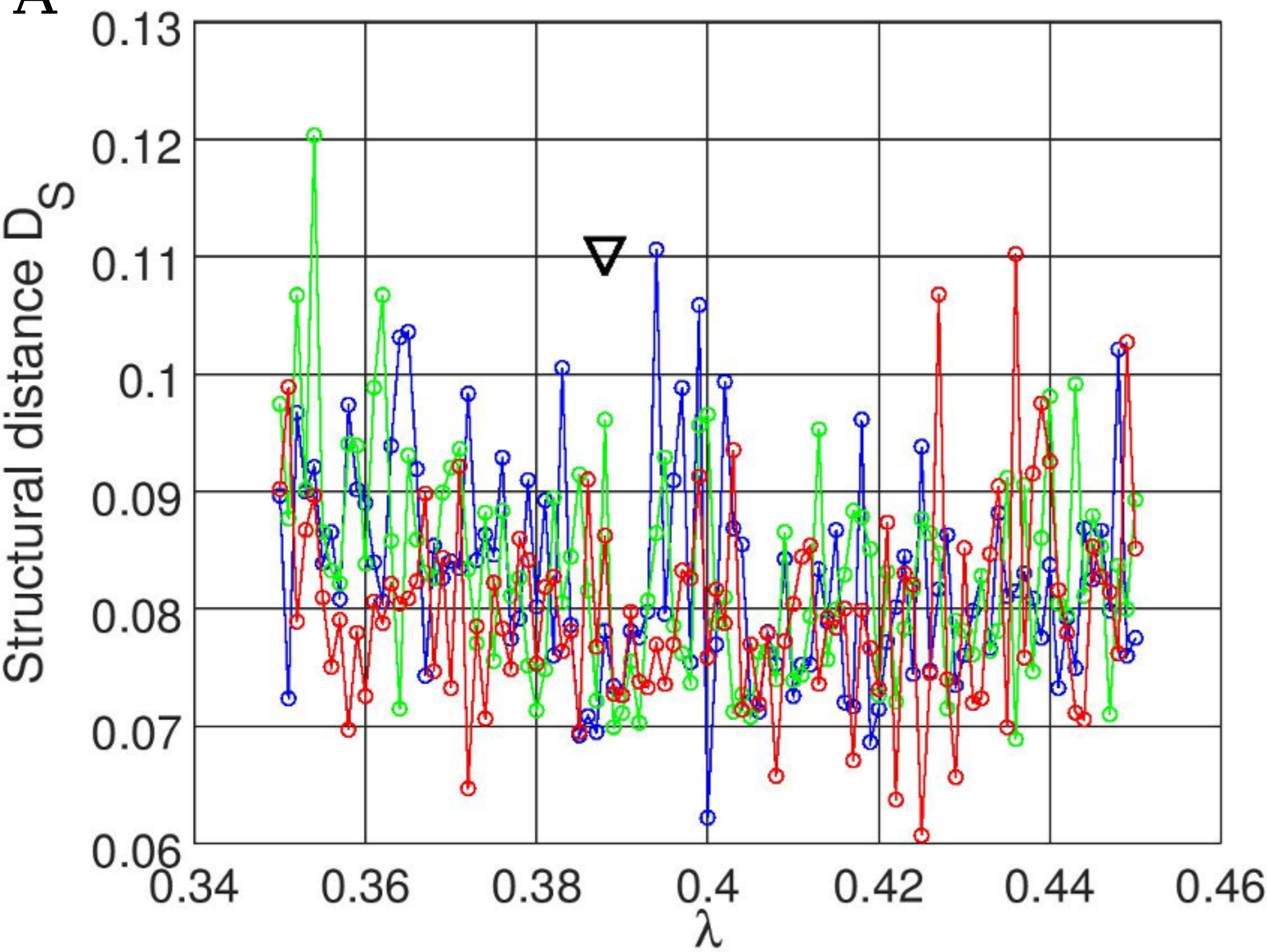
[Click here to download Figure Figure2.pdf](#)



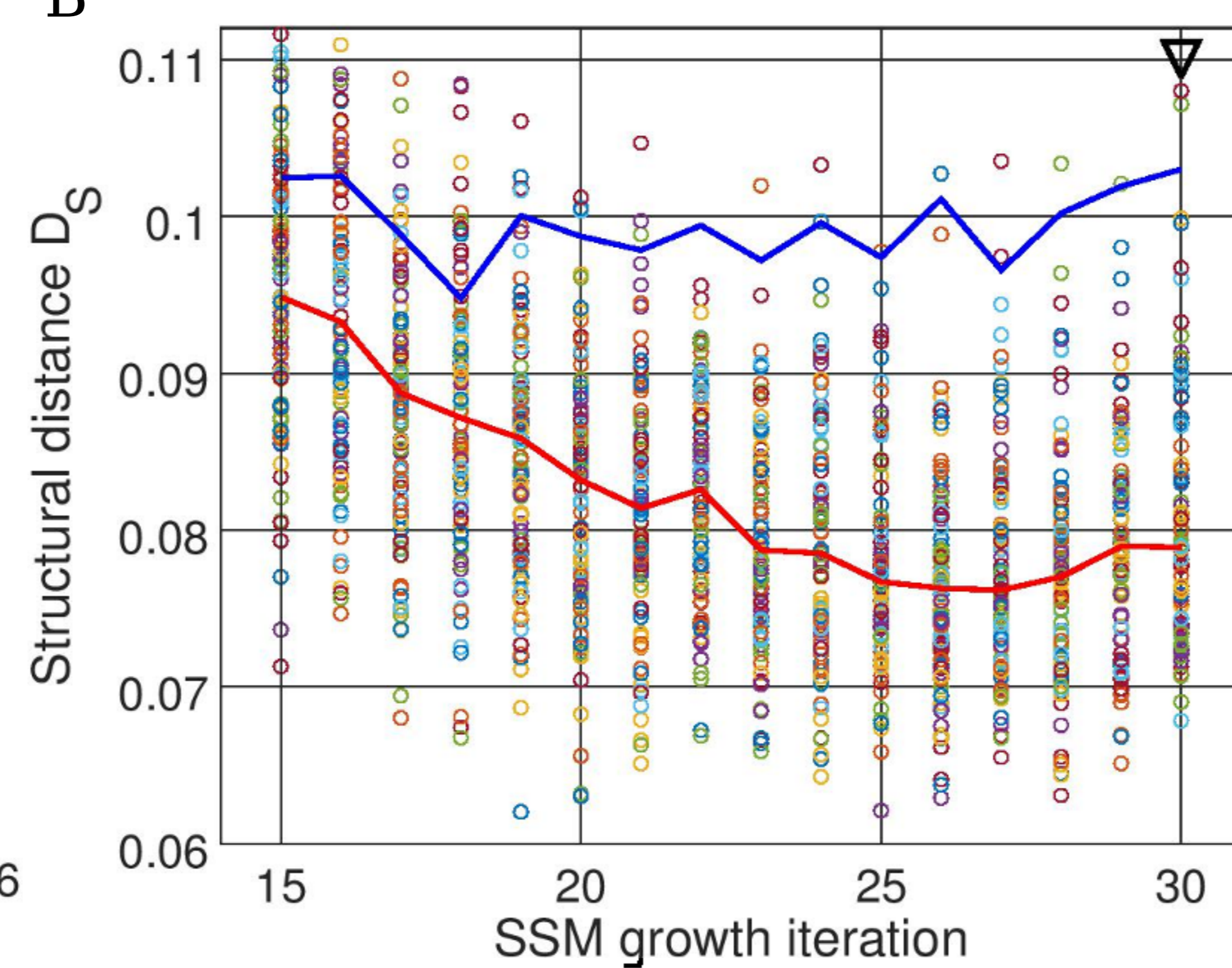




A



B



C

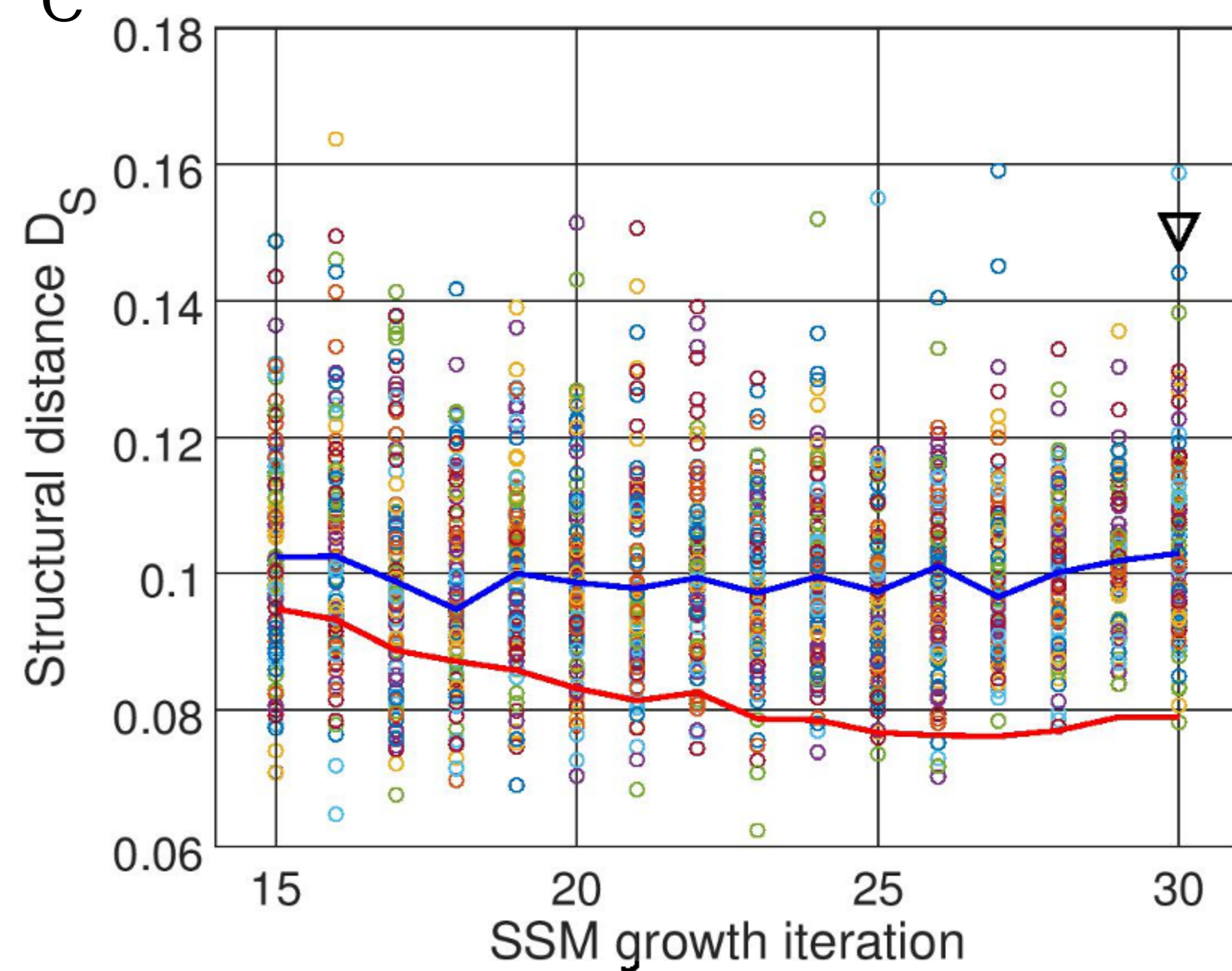


Figure 6

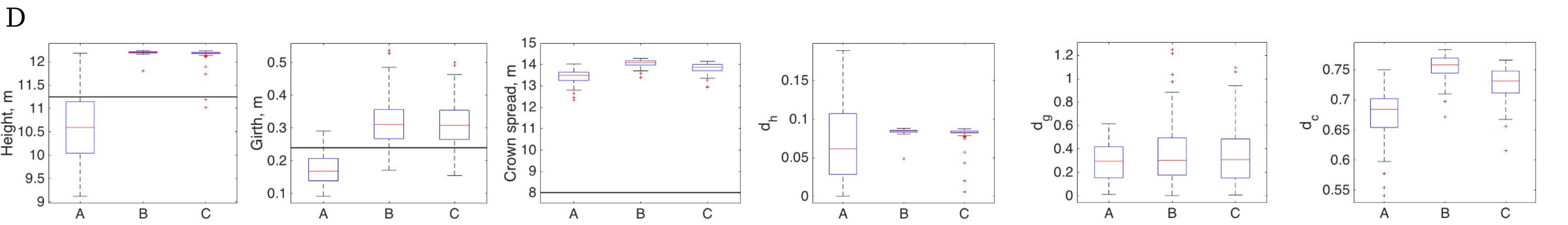
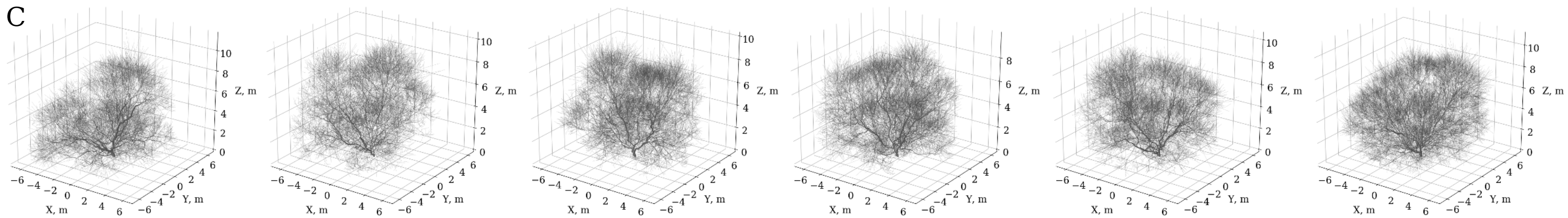
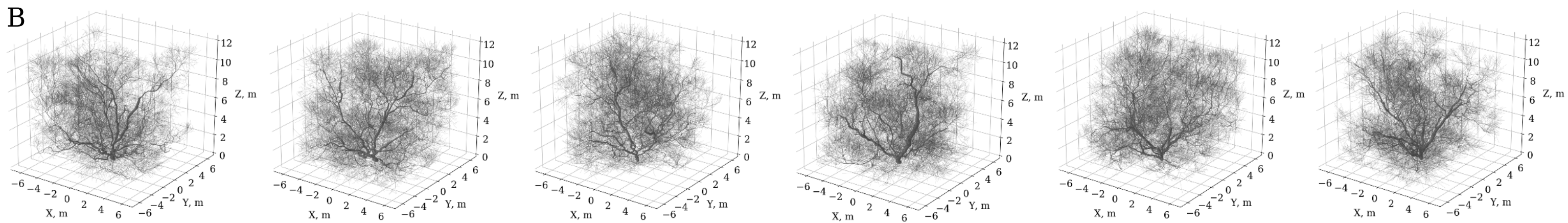
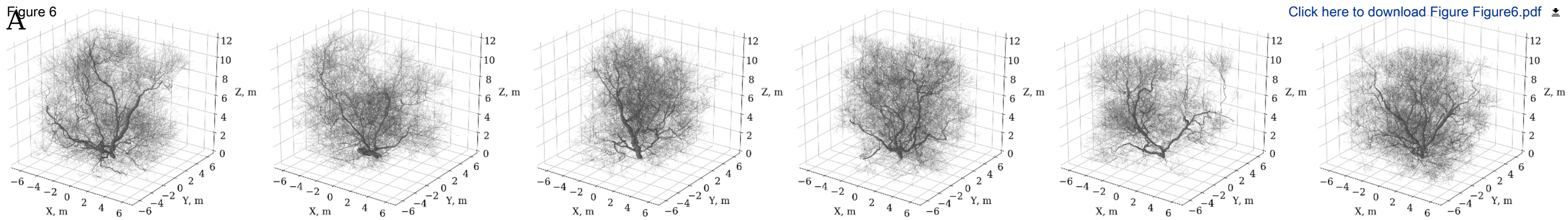
[Click here to download Figure Figure6.pdf](#)


Figure 7

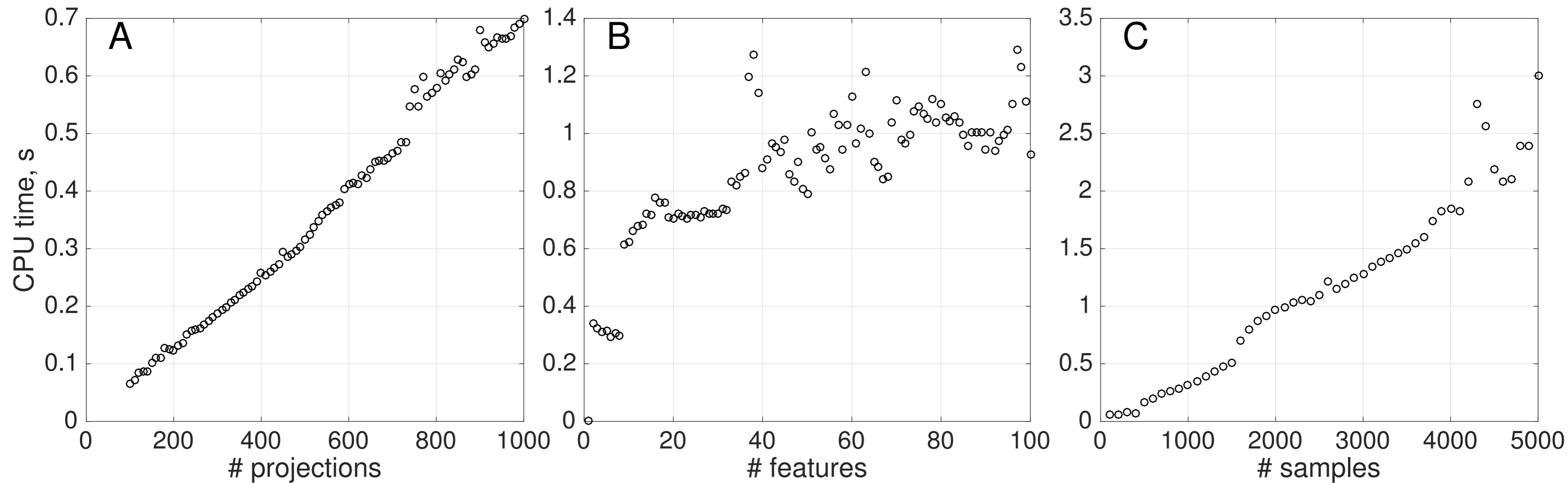
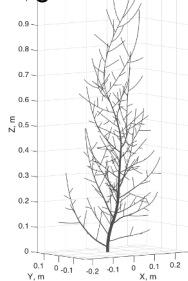
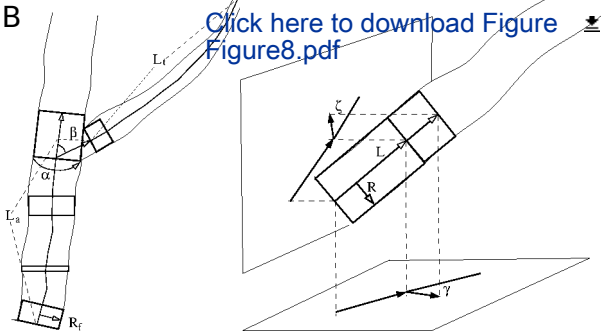
[Click here to download Figure Figure7.pdf](#)

Figure 8



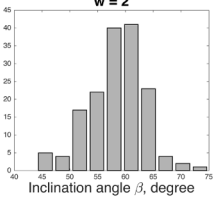
B



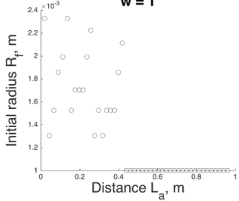
[Click here to download Figure Figure8.pdf](#)

C

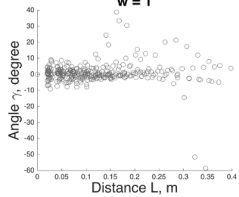
w = 2



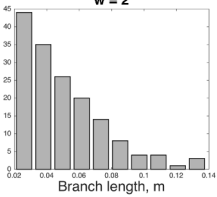
w = 1



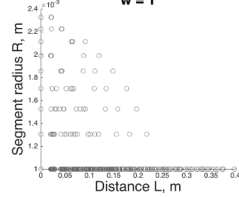
w = 1



w = 2



w = 1



w = 1

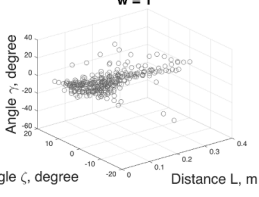
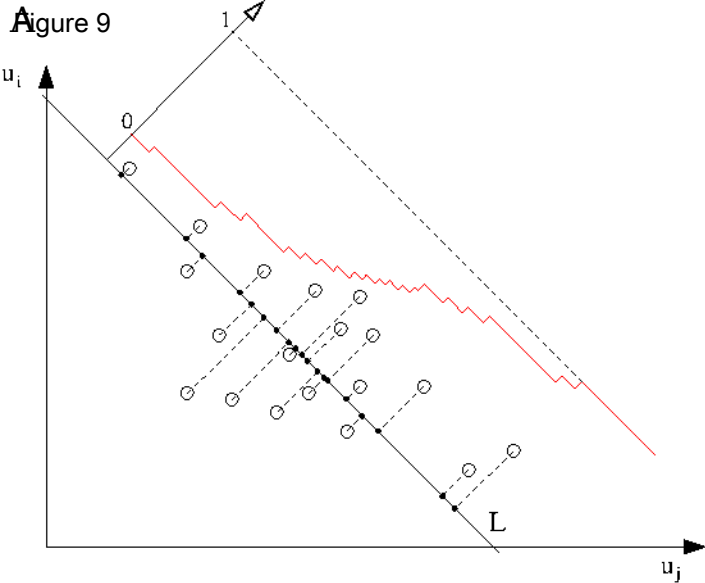


Figure 9**B** [Click here to download Figure Figure9.pdf](#) 