# Title: Characterization and Comparison of Plant Root Shapes by Stochastic Processes

## Chapter 1

I. Introduction

A. Background

i. Importance of Roots

a. Mechanical Abilities

- Supporting the aboveground structures.

b. Functional Abilities

- Uptake, storage, and transport of water and nutrients.

ii. Importance of Research

a. Under the unstable soil condition, how effective the root systems mitigate the influence of abiotic stress and maintain or improve crop production and security by phenotype modification.

b. Exploring the relationship between genetic variation and plant adaptation and plasticity to sub-optimal and heterogeneously distributed nutritions soil conditions (e.g., drought stress, insufficient phosphorus, etc.).

c. Promising application in other branching structures: river networks in geography, blood vessels in medicine, etc.

B. Summarize Current Descriptors

i. Metric

a. Basic Geometric Descriptors

- The total root length;
- The total root width;
- The area of the surface;
- The number of roots tips and branches;
- etc.

b. Compound Descriptors (Computed From the Basic Descriptors)

- Density;
- Aspect ratio;
- etc.

c. Weaknesses

- Rely on the resolution of the images[2].
- Only provide a general view of root morphology[2], which leads to a loss of information.
- Difficult to assess the spatial configuration of roots[2].
- Fail to describe the full complexity of root systems.

ii. Non-metric

a. Topological Analysis

- Methods
  - Persistent homology [11]
    - Persistence barcode shows the number of branches.
    - Persistence barcode indicates how branched roots connect along the scale of the function (e.g. geodesic distance).
    - Compare the similarity of branching structures by a pair-wise distance matrix using the bottleneck distance method.
  - Horton-Strahler index [17]
    - Categorize the topological complexity of the whole branching structure.
    - Provide a numerical measure of connectedness and complexity of the branching at each vertex by a dimensionless ratio: bifurcation ratio.
    - The range of index and the length ratio indicate the size of the branching structure.
  - Fractal Analysis
    - Measure the complexity of branching structures.
    - Measure self-similarity of branching structures [16] by fractal dimension of the root systems.
- Strengths
  - Highly complementary to geometric descriptors to characterize how individual roots are connected through branching [3].
  - Describe branching structures independent of transformation and deformation.
- Problems and Weaknesses
  - Only analyze the connectedness of branching structures, which is a portion of the complexity of plant root systems.
  - Fail to describe geometric properties of roots.
  - Fail to characterize spatial distribution.

    - Some biologically topological indices analyze the root growth qualitatively based on line-linked systems [5], but not characterize the mathematical topological properties.
    - The fractal analysis aims to describe self-similar structures, which grow by continually repeating simple growth rules [6].

C. Problem Statements

   i. Limitation of Data

      a. Large variance of length scales of root leads to the high requirement of the image resolution.

   ii. Incompleteness and Low Efficiency

      a. All the current descriptors recover partial root architectural and morphological traits independently.
      b. Roots are infinitely complicated.
      c. When exploring the latent relations between genetic variance and plant adaptation and plasticity by the multivariate data analysis, various units of the current root features cause extra efforts of researchers.
      d. The existence of functional dependence among the current descriptors can not reveal the intrinsic properties of root shape.

   iii. Incorrectness

      a. Roots are not self-similar inherently.

   iv. Lack of Computable Definitions

      a. Architecture
      b. Morphology

D. Questions

   i. Without measuring by a ruler, can we characterize the root shapes?
   ii. Can we develop novel models for analyzing the 2D root images by applied mathematics?
   iii. Based on the details provided by images:

      a. Can we characterize the full-scale roots by unitless descriptors, which do not depend on the units of the measurement?
      b. Can we compare the complicated physical properties of roots by innovative and straightforward descriptors, which will reveal the intrinsic identities, similarities, and differences between the plant roots?

II. Inspirations of Research

A. Inspiration 1: Solving Diffusion Equations

i. Kac's Ideas [10]: Can one hear the shape of a drum?

a. Summarize Kac's ideas

- Consider a two-dimensional uniform simply connected membrane bounded by a smooth convex curve (i.e. a drum).
- Apply the Dirichlet (absorbed) boundary condition.
- Calculate asymptotic expansion of the spectral function (trace function) involving eigenvalues of Laplacian operator: $\sum_n e^{-\lambda_n t}$ for small positive $t$ (hearing a discrete spectrum of pure tones produced by the drum).
- How much about the shape can be inferred from hearing all the pure tones generated by the drum?

b. Advantages of Kac's ideas

- One can deduce the geometrical properties of a plane uniform membrane from the knowledge of the distribution of eigenvalues instead of measuring by metric tools (i.e. the total area, the perimeter, the curvature, the number of holes, etc.).

c. Weakness of Kac's ideas

- Only available for the convex domain, which has a smooth or piecewise smooth boundary.
- Kac derived the asymptotic expansion indirectly from the Green's function of the corresponding diffusion equation (heat equation)[9] for an infinite wedge and then extended to smooth convex membranes by the considerations of approximating polygonal membranes [15]. Because of the difficulty with the Green's function, even for the circular membrane, this method is limited to obtain the first three terms [9].

ii. Extended Works of Kac's Idea (Spectral Function)

a. Extended Work 1 [8] [12] [13]

- Summarize the Idea
  - Consider a simply connected two-dimensional region.
  - Neumann boundary conditions.
  - Calculate asymptotic expansion of the spectral function involving eigenvalues of Laplacian operator: $\sum_n e^{-\lambda_n t}$, which converges for all positive $t$.
- Results

- In the asymptotic expansion, the coefficient of the term, which determines the perimeter ($L$) of the membrane, changes the sign compared with the Dirichlet case.
- Obtain the area of the region and its total perimeter ($L$).

b. Extended Work 2 [18]

- Summarize the Idea:
  - $\Omega$ is an arbitrary doubly-connected region in $R^2$
  - $\Omega$ is surrounded internally by a simply connected bounded domain $\Omega_1$ with a smooth boundary $\partial\Omega_1$.
  - $\Omega$ is surrounded externally by a simply connected bounded domain $\Omega_2$ with a smooth boundary $\partial\Omega_2$.
  - Determine the geometry of $\Omega$ from the asymptotic behaviour of the spectral function (trace function) for a short positive time $t$.
  - Two kinds of boundary conditions:
    - N.b.c. on $\partial\Omega_1$ and D.b.c. on $\partial\Omega_2$;
    - or D.b.c on $\partial\Omega_1$ and N.b.c on $\partial\Omega_2$;
- Results
  - The first term in the asymptotic expansion represents the area of the arbitrary doubly-connected region $\Omega$.
  - The second term in the asymptotic expansion determines the length difference of $\partial\Omega_1$ and $\partial\Omega_2$.
  - The third term in the asymptotic expansion describes the curvatures of $\partial\Omega_1$ and $\partial\Omega_2$, respectively.
  - No constant term in the asymptotic expansion compared with pure D.b.c or N.b.c. because of only one hole in the domain.

c. A Special Case of Extended Work 2 [14]

- Summarize the Idea
  - Consider an annulus
  - Explore the relationship between the spectral function (trace function) and the geometrical properties of an annulus with sets of boundary conditions.
  - Sets of boundary conditions:
    - D.b.c on inner boundary and D.b.c on outer boundary
    - D.b.c on inner boundary and N.b.c on outer boundary

* N.b.c on inner boundary and D.b.c on outer boundary
* N.b.c on inner boundary and N.b.c on outer boundary

- Results
  - The explicit asymptotic expansions of trace function with different sets of boundary conditions are different.
  - The coefficients of $t^{-1}$ determines the area of the circular annulus.
  - The coefficient of $t^{-\frac{1}{2}}$ determines the total length of the boundary of annulus.
  - The constant term is zero because the domain has only one hole.
  - The spectral function for the two-dimensional Laplacian together with D.b.c or N.b.c or mixed b.c can be written in the form $\sum_{n=0}^{\infty} C_n t^{\frac{n}{2}-1}$ as $t \to 0$.

d. Summarize the Methods Based on the Spectral Function

- The difficulty with Green's function leads to the limitation of the first three terms in the asymptotic expansion of a spectral function.
- It is impossible to simulate the spectral function of an infinite number of eigenvalues numerically.

iii. Extended Works of Kac's Idea (Heat Content) [4]

a. Summarize the Idea

- A compact smooth Riemannian manifold of dimension $m$ with $C^\infty$ boundary $\partial M$.
- Calculate the solutions for the diffusion equation with various boundary and initial conditions.
- Integrate solutions on the whole manifold, which means the total heat energy content of the manifold.
- Calculate the asymptotic series of the heat content as $t \to 0^+$
- The solutions of diffusion equation can be interpreted as probability density function of particles (or molecules) in space, evolving in time [7] so that the asymptotic series of the heat content as $t \to 0^+$ equals to the asymptotic series of survival probability as $t \to 0^+$.

b. Geometrical information of manifold described by the asymptotic expansion of survival probability.

- Volumn
- Length

- Scalar Curvature
- Mass
- etc.

c. Limitations

- Continuous space (the integration of solutions of diffusion equations).
- Continuous time ($t \to 0^+$).
- Some calculated descriptors are meaningless (e.g. scalar curvature, etc.).

iv. Disadvantages of Methods related to Kac's ideas

- It is impossible to mimic the infinite diffusion process numerically, defined in continuous time and space.
- It is impossible to calculate the infinite number of coefficients of asymptotic series of heat content numerically.
- So far, there doesn't exist any numerical simulation of the diffusion process (e.g. Kac's ideas, heat content, etc.) for describing the shape of 2D objects.

B. Inspirations 2: Dimensional Analysis

i. Reduce the number of variables in the problem without losing any information and have a more appropriate interpretation of the problem [1].

ii. Illustrate the intrinsic properties of the phenomenon by fewer dimensionless quantities or a complete set of dimensionless products of given variables, which do not depend on arbitrarily chosen basic units of measurement.

iii. Test the validness of the model by dimensional homogeneity.

iv. Furnish valuable qualitative indications, the quantity information, and functional dependence of the features.

v. Find self-similarity of problems.

III. Research Ideas

A. Statements

i. Stochastic processes, defined in the finite discrete space and time, are idealized and innovative mathematical models to characterize the root shape without measuring all the details, provided by the images, by rulers.

ii. We can generate novel dimensionless descriptors in stochastic processes to reveal and compare the intrinsic physical properties of the full-scale root in 2D images.

# Chapter 2

IV. Stochastic Process 1 — Diffusion

A. Background of Diffusion

i. The analytical asymptotic infinite series of heat content defined in continuous time and space as $t \to 0^+$.

ii. As the step size of particles in the random walk model tends to 0 and the number of steps increases proportionally, random walk converges to a Wiener process (Brownian motion) in an appropriate sense.

a. Unbiased random walk model in the finite discrete space and time

- Definition: particles have the same probabilities of walking in random directions with a fixed step length.
- Model parameters in 2D images
  - Boundary conditions of domain and target.
  - Initial condition for particles.
  - Directions of particles for each step.
  - Step length of particles or spacing grid of the domain.
  - Required number of particles.

iii. Dimensional Analysis

B. Survival Analysis

i. Significance

a. Defined in continuous time and space, the mathematical meaning of the survival probability of individuals in the diffusion process is as same as the heat content.

ii. Survival Probability

a. Analytical Calculation

- Definition
- Calculated by integrating the solutions of diffusion equations.
- Derive the geometrical information of the boundary from analytical asymptotic infinite series.

- Expect the features of survival curves based on the analytical expression.
- Note : Since it is impossible to calculate the infinite asymptotic series of survival probability defined in continuous time and space, there only exists a general format of the asymptotic series.

b. Numerical Estimation and Comparison
   - Kaplan-Meier estimator.
   - Properties.
   - Statistical significant tests of $N \geq 2$ groups of survival curves.
     - Wasserstein or Kantorovich–Rubinstein metric or distance
   - Statistical significant tests for $n \geq 2$ survival curves.
     - General methods
       - Log rank test: measures and reports on whether two intensity processes are different.
       - Multiple pairwise log rank test: $n \geq 2$ survival curves.
       - Note: log rank test may be not reliable
         - If there are many time intervals, the test will be arduous.
         - It requires an important assumption: the proportional hazard ratio is a constant.
         - If the survival curves cross, the logrank test will give an inaccurate assessment of differences
     - Specific tests for right-censoring data
       - Gehan-Breslow-Wilcoxon test
         - Doesn't require the assumption of the proportional hazards to be met.
         - Doesn't work well for crossing survival curves.
         - Give more weight to the early failures of survival curves.
       - Prentice test (Prentice modified Wilcohon test or Peto-Peto-Prentice test)
         - Gives more emphasis on the earlier event times.
         - The proportional hazards assumption is violated.
         - Censoring rates are low and censoring distributions of groups are equal.
         - Doesn't work well for crossing survival curves.
       - Tarone-Ware test

- Works well for crossing curves.
- Emphasize more on the failures that happened somewhere in the middle of the course.
- Not limited to the number of groups to be applied for.

iii. Analytical Mean First-passage Time

a. Definition.

b. Calculated from the survival probability.

C. Regular Simple Shape: An Annulus

i. Analytical Solutions of Diffusion Equation

a. The Meaning of Solutions

- ß is a probability density function, which gives the value of diffusing particles at $\left(\hat{r}, \theta\right)$ at time $\tau$.

b. Conditions of Partial Differential Equations

- Initial condition: uniform distribution
- Boundary conditions
  - D.b.c (absorbed) for target (inner) boundary
  - N.b.c (reflecting) for escape (exterior) boundary
- The reasons of boundary conditions
  - If boundary conditions are inverse, the survival probability will indicate the geometrical properties of the escape boundary.
  - If both boundaries have N.b.c, the survival probability will be a constant, which will not be helpful to characterize the shape of target boundary.

c. Methods of Solving Diffusion Equations

- Separation of variables method.
- Dimensional analysis.

d. Analytical Survival Analysis

- MFPT reveals geometrical information of the annulus.

ii. Unbiased Random Walk Model Design

a. Model Parameters

- D.b.c (absorbed) for target (inner) boundary
- N.b.c (reflecting) for escape (exterior) boundary

- Initial condition: uniform distribution
- Directions: up, down, left, right
- Step Length: 1 pixel
- Spacing grid
  - Shannon sampling theorem
  - Dimensional analysis
- Required number of particles
  - The mean of the mean first-passage time
  - The law of large numbers
  - Chebyshev's inequality

b. Particles' motions

iii. Test Numerical Results

a. After converting the numerical steps into dimensionless time, compare the analytical and numerical survival curves.

b. Compare the analytical and estimated mean of the mean first-passage time.

iv. Conclusion

a. Survival analysis of discrete and finite numerical random walk model is innovative and helpful to mimic the continuous and infinite analytical diffusion process for characterizing the intrinsic features of the annulus.

b. Influence factors of particles' motion

- Step length.
- Boundary conditions of the domain.
- The size of domain.
- Physical properties of the targets.

D. Artificial Images: Branching Structures

i. Image Description

a. The size of images is identical.

b. The area of the objects in the pictures (the number of white pixels) is identical.

- Group 1: Objects have random locations.
- Group 2: Corresponding to group 1, shapes are different.
- Group 3: Objects located in the center of the images.

ii. Model Design 1

a. Model Parameters

- D.b.c on target.
- N.b.c on top and bottom boundaries of the domain.
- P.b.c on the left and right boundaries of the domain (mimic infinite arrays of plant roots in the soil).
- Initial condition: uniform distribution.
- Directions: $\theta \in \big[0, 2\pi\big)$.
- Step length doesn't depend on the spacing grid of the image $l \in \big(0, 1]$.
- The required number of particles.

b. Particles' motions

- Particles are distributed uniformly in the whole artificial image at step 0.
- If the particles are in the target or touch the target boundary, they will stop moving. If not, they will continue to walk.
  - If particles touch the reflecting side, they will walk in the opposite direction.
  - If particles reach the periodic boundary, they will appear on the opposite side.
- Particles will stop when they touch the boundary of the target. Store the total number of steps as the first hitting time.

c. Survival Curves' Description

- Analyze the distributions of initial and stop positions of particles
  - The distributions visualize the behaviours of particles in the model.
  - Full-scale geometrical features of the shape.
  - Spatial distribution of the target shape.
- The estimated survival probability as $t = 0$ is as similiar as the area ratio $\frac{A_{target}}{A_{domain}}$.

d. Statistical Tests of Survival Curves

e. Conclusion

- Survival curves visualize the asymptotic expansion of heat content.
- Short-time survival curves describe the full-scale geometrical properties and spatial configuration of objects in the images.

- Survival curves are not statistically significant for two same objects whose locations are different.
- Survival curves are statistically significant for two objects whose differences are tiny.

iii. Model Design 2:

a. Right-censoring Data in Survival Analysis

- Definition
- Data format
  - Set a maximum steps: all the hitting times are less than a maximum value.
    - Only the short time behaviours of particles indicate the geometrical information of the target object.
    - There are specific statistical tests for right-censoring data in survival analysis.
    - Improve the efficiency of numerical simulation.
  - Duration of each particle: the total number of steps.
  - Event: label the particle as 0 when the study finished, but the event of the interest was not observed; label the particle as 1 when it reached the target boundary.

b. Model Parameters

- Initial condition: uniform distribution.
- Directions: $\theta \in \big[0, 2\pi\big)$.
- Step length doesn't depend on the spacing grid of the image $l \in \big(0, 1\big]$.
- The required number of particles.

c. Particles' motions

- Particles are distributed uniformly in the whole artificial image at step 0.
- If the particles are in the target or touch the target boundary, they will stop moving and be labelled as 1. If not, they will continue to walk.
  - If particles touch the boundary of the target and the total number of steps is less than the threshold, they stop walking and be labelled as 1.
  - If the total number of steps of particles is more than the threshold or particles reach the boundaries of the domain, label them as 0.

- Store the total number of steps of each particle as the first hitting time.
- Calculate the distance between the initial and stop point as the displacement for each particle.
- Calculate the distance between the initial and farthest point as the maximum displacement for each particle.

d. Statistical Significance of Survival Curves of Right-censoring Data

- First hitting time
- Displacement
- Maximum Displacement

e. Analyze the distributions of initial and stop positions of particles

- First hitting time
- Displacement
- Maximum Displacement

f. Conclusion

- Survival curves of displacement and maximum displacement of particles can also visualize particles' behaviours and geometrical and spatial features of the 2D objects.
- Statistical tests of survival analysis are also reliable for comparing the censored survival data and short-time behaviours of curves.
- It is crucial to choose a relative reasonable statistical test for survival curves.

E. Real Root Images

i. Image Description

a. The size of images is same.

b. Compared with unprocessed images, processed sorghum images are clean and segmented into binary form, where background pixels are 0, and foreground pixels are 255.

c. Two genotypes: PI533752 and PI561073.

d. Treatments: high P and low P for 7, 10, and 14 days with five repeated experiments.

ii. Model Parameters

a. Initial condition: uniform distribution.

b. Directions: $\theta \in \big[0, 2\pi\big)$.

c. Step length doesn't depend on the spacing grid of the image $l \in \big(0, 1\big]$.

d. Required number of particles for each image

iii. Statistical Significance of Two Groups of Survival Curves

F. Conclusion

i. Survival analysis of the unbiased random walk model develops a reproducible and simple curve as a novel unitless descriptor for representing and comparing the full-scale physical properties of the 2D root shapes.

a. Compare two survival curves can reveal the phenotypic variance of two root systems since each 2D root image has a specific survival curve.

b. Compare two groups of survival curves, which represent two groups of plant root systems with different genotypes, can statistically unveil the relationship between genetic variation and plant adaptation and plasticity.

# References


[1] P Bag and N Zealand. "Simplifying regression models using dimensional analysis, no. 1981". In: (2001).

[2] Mathilde Balduzzi et al. "Reshaping plant biology: qualitative and quantitative descriptors for plant morphology". In: *Frontiers in plant science* 8 (2017), p. 117.

[3] Benjamin M Delory et al. "archiDART v3. 0: A new data analysis pipeline allowing the topological analysis of plant root systems". In: *F1000Research* 7 (2018).

[4] S Desjardins and P Gilkey. "Heat content asymptotics for operators of Laplace type with Neumann boundary conditions". In: *Mathematische Zeitschrift* 215.1 (1994), pp. 251–268.

[5] AH Fitter. "The topology and geometry of plant root systems: influence of watering rate on root system topology in Trifolium pratense". In: *Annals of Botany* 58.1 (1986), pp. 91–101.

[6] AH Fitter and TR Stickland. "Fractal characterization of root system architecture". In: *Functional Ecology* (1992), pp. 632–635.

[7] Rudolf Gorenflo et al. "Discrete random walk models for space–time fractional diffusion". In: *Chemical physics* 284.1-2 (2002), pp. 521–541.

[8] HPW Gottlieb. "Eigenvalues of the Laplacian with Neumann boundary conditions". In: *The ANZIAM Journal* 26.3 (1985), pp. 293–309.

[9] HPW Gottlieb. "Hearing the shape of an annular drum". In: *The ANZIAM Journal* 24.4 (1983), pp. 435–438.

[10] Mark Kac. "Can one hear the shape of a drum?" In: *The american mathematical monthly* 73.4P2 (1966), pp. 1–23.

[11] Mao Li et al. "Persistent homology and the branching topologies of plants". In: *American journal of botany* 104.3 (2017), pp. 349–353.

[12] Åke Pleijel. "A study of certain Green's functions with applications in the theory of vibrating membranes". In: *Arkiv för Matematik* 2.6 (1954), pp. 553–569.

[13] BD Sleeman. "The inverse problem of acoustic scattering". In: *IMA Journal of Applied Mathematics* 29.2 (1982), pp. 113–142.

[14] BD Sleeman and EME Zayed. "Trace formulae for the eigenvalues of the Laplacian". In: *Zeitschrift für angewandte Mathematik und Physik ZAMP* 35.1 (1984), pp. 106–115.

[15] K Stewartson and RT Waechter. "On hearing the shape of a drum: further results". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 69. 2. Cambridge University Press. 1971, pp. 353–363.

[16] JIRO TATSUMI, AKIRA YAMAUCHI, and YASUHIRO KONO. "Fractal analysis of plant root systems". In: *Annals of Botany* 64.5 (1989), pp. 499–503.

[17] Zoltán Toroczkai. "Topological classification of binary trees using the Horton-Strahler index". In: *Physical Review E* 65.1 (2001), p. 016130.

[18] EME Zayed. "Heat equation for an arbitrary doubly-connected region inR 2 with mixed boundary conditions". In: *Zeitschrift für angewandte Mathematik und Physik ZAMP* 40.3 (1989), pp. 339–355.