Horton self-similarity and coalescent trees

Horton self-similarity of Kingman’s coalescent

Yevgeniy Kovchegov
Oregon State University

(joint work with Ilya Zaliapin of U. Nevada)
Introduction.

When studying the tree graphs associated with random structures one often aims at discovering a particular symmetry or a consistent pattern such as self-similarity. There exist two important types of tree self-similarity related to the Horton-Strahler ordering and Tokunaga indexing schemes for tree branches.

The **Horton-Strahler indexing** assigns orders to the tree branches according to their relative importance in the hierarchy.

- Introduced in hydrology in the 1950s to describe the dendritic structure of river networks.
- Applications: ranking river tributaries, analysis of brain structure, designing optimal computer codes, etc.
Self-similar trees

- A two-parametric class of *Tokunaga self-similar trees* closely approximates a surprising variety of trees in observed and modeled systems [Tokunaga, 1978; Peckham, 1995; Newman et al., 1997; Zanardo et al., 2013]

- Tokunaga self-similarity implies *Horton laws*, heavily used in hydrology since the 1950-s

- Horton laws can be interpreted as a power-law distribution of system element sizes, and hence are relevant to many hierarchical systems
Horton-Strahler ordering.

- Horton and Tokunaga laws are based on *Horton-Strahler orders* that measure “importance” of tree branches within the hierarchy.

- In a perfect binary tree (all leaves having the same depth) the orders are proportional to depth.

- How to assign orders in a non-perfect tree?
Horton-Strahler ordering.

Rooted tree
Horton self-similarity and coalescent trees

Horton-Strahler ordering.

Rooted tree
Horton-Strahler ordering.

Rooted tree
Horton self-similarity and coalescent trees

Horton-Strahler ordering.

Rooted tree

Vertex b is parent of vertex a; vertex a is a child of vertex b
Horton-Strahler ordering.

Rooted tree

Vertex b is parent of vertex a; vertex a is a child of vertex b
Horton-Strahler order (via pruning).

- Pruning $R(T)$ of a finite tree $T$ cuts the leaves and degree-2 chains connected to leaves.

- Nodes cut at $k$-th pruning, $R^{k-1}(T) \setminus R^k(T)$, have order $k$, $k \geq 1$.

- A chain of the same order vertices is called *branch*.

- $N_k$ denotes the number of *branches* of order $k$ in a finite tree $T$. 
Horton-Strahler order (via pruning).

- *Pruning* $\mathcal{R}(T)$ of a finite tree $T$ cuts the leaves and degree-2 chains connected to leaves.

- Nodes cut at $k$-th pruning, $\mathcal{R}^{k-1}(T) \setminus \mathcal{R}^k(T)$, have order $k$, $k \geq 1$.

- A chain of the same order vertices is called *branch*.

- $N_k$ denotes the number of *branches* of order $k$ in a finite tree $T$
Horton self-similarity and coalescent trees

Horton-Strahler order (via pruning).

- *Pruning* $\mathcal{R}(T)$ of a finite tree $T$ cuts the leaves and degree-2 chains connected to leaves.

- Nodes cut at $k$-th pruning, $\mathcal{R}^{k-1}(T) \setminus \mathcal{R}^k(T)$, have order $k$, $k \geq 1$.

- A chain of the same order vertices is called *branch*.

- $N_k$ denotes the number of *branches* of order $k$ in a finite tree $T$.
Horton-Strahler order (via pruning).

- Pruning $\mathcal{R}(T)$ of a finite tree $T$ cuts the leaves and degree-2 chains connected to leaves.

- Nodes cut at $k$-th pruning, $\mathcal{R}^{k-1}(T) \setminus \mathcal{R}^k(T)$, have order $k$, $k \geq 1$.

- A chain of the same order vertices is called branch.

- $N_k$ denotes the number of branches of order $k$ in a finite tree $T$.
Horton self-similarity and coalescent trees

Horton-Strahler order (via pruning).

- Pruning $\mathcal{R}(T)$ of a finite tree $T$ cuts the leaves and degree-2 chains connected to leaves.

- Nodes cut at $k$-th pruning, $\mathcal{R}^{k-1}(T) \setminus \mathcal{R}^k(T)$, have order $k$, $k \geq 1$.

- A chain of the same order vertices is called branch.

- $N_k$ denotes the number of branches of order $k$ in a finite tree $T$
Horton-Strahler ordering and Tokunaga indexing.

Example: (a) Horton-Strahler ordering
(b) Tokunaga indexing.

Two order-2 branches are depicted by heavy lines in both panels. The Horton-Strahler orders refer, interchangeably, to the tree nodes or to their parent links. The Tokunaga indices refer to entire branches, and not to individual vertices.
Horton-Strahler ordering.

The **Horton-Strahler ordering** of the vertices of a finite rooted labeled binary tree is performed in a hierarchical fashion, from leaves to the root:

(i) each leaf has order $r(\text{leaf}) = 1$;

(ii) when both children, $c_1, c_2$, of a parent vertex $p$ have the same order $r$, the vertex $p$ is assigned order $r(p) = r + 1$;

(iii) when two children of vertex $p$ have different orders, the vertex $p$ is assigned the higher order of the two.
Tokunaga indexing.

- Let $\tau_{ij}^{(k)}$, $1 \leq k \leq N_j$, $1 \leq i < j \leq \Omega$, denote the number of branches of order $i$ that join the non-terminal vertices of the $k$-th branch of order $j$.

- Let $N_{ij}$ be the total number of instances when an order-$i$ branch merges an order-$j$ branch:

$$N_{ij} = \sum_{k} \tau_{ij}^{(k)}, i < j$$

- The Tokunaga index $T_{ij}$ is the average number of order-$i$ branches that join an order-$j$ branch:

$$T_{ij} = \frac{N_{ij}}{N_j}$$
Tokunaga indexing.

- Let $\tau_{ij}^{(k)}$, $1 \leq k \leq N_j$, $1 \leq i < j \leq \Omega$, denote the number of branches of order $i$ that join the non-terminal vertices of the $k$-th branch of order $j$.

- Let $N_{ij}$ be the total number of instances when an order-$i$ branch merges an order-$j$ branch:
  $$N_{ij} = \sum_k \tau_{ij}^{(k)}, i < j$$

- The Tokunaga index $T_{ij}$ is the average number of order-$i$ branches that join an order-$j$ branch:
  $$T_{ij} = \frac{N_{ij}}{N_j}$$
Tokunaga indexing.

- Let $\tau_{ij}^{(k)}$, $1 \leq k \leq N_j$, $1 \leq i < j \leq \Omega$, denote the number of branches of order $i$ that join the non-terminal vertices of the $k$-th branch of order $j$.

- Let $N_{ij}$ be the total number of instances when an order-$i$ branch merges an order-$j$ branch:
  \[ N_{ij} = \sum_k \tau_{ij}^{(k)}, i < j \]

- The Tokunaga index $T_{ij}$ is the average number of order-$i$ branches that join an order-$j$ branch:
  \[ T_{ij} = \frac{N_{ij}}{N_j} \]
Tree self-similarity

**Definition 1** (Self-similarity). A random tree $T$ of order $\Omega$ is self-similar if

$$E \left[ \tau_{i(i+k)}^{(j)} \right] =: T_k$$

for $1 \leq j \leq N_{i+k}$, $2 \leq i + k \leq \Omega$.

**Definition 2** (Tokunaga self-similarity). A random self-similar tree is *Tokunaga self-similar* if

$$T_{k+1}/T_k = c \quad \Leftrightarrow \quad T_k = a c^{k-1} \quad a, c > 0, \ 1 \leq k \leq \Omega - 1.$$
**Tree self-similarity**

The matrix of Tokunaga indices

\[
T = \begin{bmatrix}
T_{12} & T_{13} & T_{14} & \cdots & T_{1\Omega} \\
0 & T_{23} & T_{24} & \cdots & T_{2\Omega} \\
0 & 0 & T_{34} & \cdots & T_{3\Omega} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & T_{\Omega-1\Omega}
\end{bmatrix}
\]

becomes a Toeplitz matrix for a self-similar tree:

\[
T_{SS} = \begin{bmatrix}
T_1 & T_2 & T_3 & \cdots & T_{\Omega-1} \\
0 & T_1 & T_2 & \cdots & T_{\Omega-2} \\
0 & 0 & T_1 & \cdots & T_{\Omega-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & T_1
\end{bmatrix}; \quad T_{TSS} = \begin{bmatrix}
a & ac & ac^2 & \cdots & ac^{\Omega-2} \\
0 & a & ac & \cdots & ac^{\Omega-3} \\
0 & 0 & a & \cdots & ac^{\Omega-4} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & a
\end{bmatrix}
\]
Horton self-similarity.

We say that a sequence of probability laws \( \{ \mathcal{P}_N \}_{N \in \mathbb{N}} \) has **well-defined asymptotic Horton-Strahler orders** if for each \( k \in \mathbb{N} \), random variables

\[
\frac{\mathcal{N}_k^{(\mathcal{P}_N)}}{N} \xrightarrow{\text{in probability}} \mathcal{N}_k \quad \text{as} \quad N \to \infty,
\]

where quantity \( \mathcal{N}_k \) is called the **asymptotic ratio** of the branches of order \( k \).

The notion of **Horton self-similarity** characterizes the cases when the sequence \( \mathcal{N}_k \) decreases in a regular geometric fashion with \( k \) going to infinity. Informally,

\[
\mathcal{N}_k \approx N_0 \cdot R^{-k}
\]
Horton self-similarity.

The notion of Horton self-similarity characterizes the cases when the sequence $N_k$ decreases in a regular geometric fashion with $k$ going to infinity. Informally, $N_k \asymp N_0 \cdot R^{-k}$. Formally, we define three types of Horton self-similarity.

A sequence $\{P_N\}_{N \in \mathbb{N}}$ of probability laws over binary trees with well-defined asymptotic Horton-Strahler orders is said to obey a Horton self-similarity law if and only if at least one of the following limits exists and is finite and positive:

(a) root law: $\lim_{k \to \infty} \left( \frac{N_k}{N_{k+1}} \right)^{\frac{1}{k}} = R > 0$,

(b) ratio law: $\lim_{k \to \infty} \frac{N_k}{N_{k+1}} = R > 0$,

(c) geometric law: $\lim_{k \to \infty} N_k \cdot R^k = N_0 > 0$.

The constant $R$ is called the Horton exponent.
Background.

- A classical model that exhibits Horton and Tokunaga self-similarity is critical binary Galton-Watson tree (Burd, Waymire, and Winn, 2000). This model has $R = 4$ and $(a, c) = (1, 2)$.

**Theorem [Shreve, 1969; Burd et al., 2000].** A critical binary Galton-Watson tree is Tokunaga self-similar with $(a, c) = (1, 2)$, that is

$$T_k = 2^{k-1} \quad \text{and} \quad R = 4.$$

**Theorem [Burd et al., 2000].**

1. Let $P_{GW}(p_k)$ denote the Galton-Watson distribution on the space of finite trees with branching probabilities $p_k$, $k = 0, 1, \ldots$. A tree $T \sim P_{GW}(p_k)$ is self-similar if and only if \{p_k\} is the critical binary distribution $p_0 = p_2 = 1/2$.

2. Any critical Galton-Watson tree $T$, $\sum k p_k = 1$, converges to the binary critical tree under the operation of pruning, $R^n(T)$, $n \to \infty$. 
Background.

- Peckham’95: high-precision extraction of river channels for Kentucky River, Kentucky and Powder River, Wyoming.

Reported Horton exponents and Tokunaga parameters: $R \approx 4.6$ and $(a, c) \approx (1.2, 2.5)$.

River networks: Shreve 1966, 1969; Tokunaga, 1978; Peckham, 1995; Burd et al., 2000; Z et al., 2009; Zanardo et al., 2013
Background.

- Beyond river networks: botanical leaves, diffusion limited aggregation, two dimensional site percolation, nearest-neighbor clustering in Euclidean spaces, a general hierarchical coagulation model of Gabrielov introduced in the framework of self-organized criticality, etc.
Background.


This expands the class of Horton and Tokunaga self-similar processes beyond the critical binary Galton-Watson branching, since the tree representation of Markov chains in general is not equivalent to the Galton-Watson process.
Level-set tree of a function.

(a) Function $X_t$  
(b) Tree $\text{level}(X)$

Function $X_t$ (panel a) with a finite number of local extrema and its level-set tree $\text{level}(X)$ (panel b).
Pruning of time series

Proposition [Zaliapin and K., 2012]: The transition from a time series $X_k$ to the time series $X_k^{(1)}$ of its local minima corresponds to the pruning of the level-set tree level($X$). Formally,

$$\text{level} \left( X^{(m)} \right) = \mathcal{R}^m \left( \text{level}(X) \right), \forall m \geq 1,$$

where $X^{(m)}$ is obtained from $X$ by iteratively taking local minima $m$ times (i.e., local minima of local minima and so on.)
Horton and Tokunaga self-similarity for Markov chains

Let $X_k$, $k = 1, \ldots, N$ be a symmetric homogeneous Markov chain and $T = \text{shape}(\text{level}(X))$ be the combinatorial level set tree of $X_k$.

**Theorem [Zaliapin and K., 2012].**

1. Tree $T$ is Tokunaga self-similar with parameters $(a, c) = (1, 2)$:

$$E \left[ \tau_{i(i+k)}^{(j)} \right] =: T_k = 2^{k-1},$$

and geometric-Horton self-similar, asymptotically in $N$, with $R = 4$.

2. Accordingly, a combinatorial level-set tree for regular Brownian motion is Tokunaga and Horton self-similar, with $(a, c) = (1, 2)$, and $R = 4$. 
Horton and Tokunaga self-similarity for fractional Brownian motions

Conjecture [Zaliapin and K., 2012]. The tree tree \((B^H)\) of a fractional Brownian motion \(B^H_t, t \in [0,1]\) with the Hurst index \(0 < H < 1\) is Tokunaga self-similar with \(T_{i(i+k)} = T_k = c^{k-1}, c = 2H + 1, i, k \geq 1.\)
Background.

- K. and Zaliapin, 2015: established the root-Horton law for the Kingman’s coalescent. Showed that the tree that describes a Kingman’s coalescent is combinatorially equivalent to the level-set tree of a white noise.

Perform numerical experiments that suggest that the Kingman’s coalescent, and hence the level-set tree of a white noise, are Horton self-similar in a regular stronger sense as well as asymptotically Tokunaga self-similar.
Finite coalescent process via a collision kernel.

[Markus, 1968; Lushnikov, 1978; Aldous, 1999; Pitman, 2006]

- The process starts at $t = 0$ with $N$ particles (clusters) of mass one.

- The cluster formation is governed by a collision rate kernel

  \[ K(i, j) = K(j, i) > 0, \]

  \[ 1 \leq i, j \leq N - 1. \] Specifically, a pair of clusters with masses $i$ and $j$ coalesces at the rate $K(i, j)/N$, independently of the other pairs, to form a new cluster of mass $i + j$.

- The process continues until there is a single cluster of mass $N$. 
Kingman's $N$-coalescent process.

The best studied coalescent processes (as $N \to \infty$) are:

- Kingman’s coalescent: $K(x, y) \equiv 1$
- Additive coalescent: $K(x, y) = x + y$
- Multiplicative coalescent: $K(x, y) = xy$

Coalescent tree.

A merger history of Kingman’s $N$-coalescent process can be naturally described by a time-oriented binary tree $T_{K}^{(N)}$ constructed as follows.

Start with $N$ leaves that represent the initial $N$ particles and have time mark $t = 0$. When two clusters coalesce (a transition occurs), merge the corresponding vertices to form an internal vertex with a time mark of the coalescent.

The final coalescence forms the tree root.

The resulting time-oriented tree represents the history of the process. It is readily seen that there is one-to-one map from the trajectories of an $N$-coalescence process onto the time-oriented trees with $N$ leaves.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Coalescent tree.
Results.

In Zaliapin and K. 2015, we consider the asymptotic proportion

\[ N_k = \lim_{N \to \infty} \frac{N_k}{N} \]

of the number \( N_k \) of branches of Horton-Strahler order \( k \) in Kingman's \( N \)-coalescent process with constant collision kernel.

We have a construction that allows one to interpret \( N_k \) also as the proportion of branches of order \( k \) in the infinite tree that corresponds to the Kingman's coalescent.

We show that

\[ N_k = \frac{1}{2} \int_0^\infty g_k^2(x) \, dx, \]

where the sequence \( g_k(x) \) solves:

\[ g'_{k+1}(x) - \frac{g_k^2(x)}{2} + g_k(x)g_{k+1}(x) = 0, \quad x \geq 0 \]

with \( g_1(x) = 2/(x + 2) \), \( g_k(0) = 0 \) for \( k \geq 2 \).
**Results.**

Equivalent relation:

\[ \mathcal{N}_k = \int_0^1 [1 - (1 - x) h_{k-1}(x)]^2 dx, \]

where

\[ h'_{k+1}(x) + h^2_k(x) - 2h_k(x)h_{k+1}(x) = 0, \quad x \in [0, 1] \]

with \( h_0 \equiv 0, \; h_1 \equiv 1, \) and \( h_k(0) = 1 \) for \( k \geq 1. \)

**Theorem.** The asymptotic Horton ratios \( \mathcal{N}_k \) exist and finite and satisfy the convergence \( \lim_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = R \) with \( 2 \leq R \leq 4. \)

**Conjecture.** The tree associated with Kingman’s coalescent process is Horton self-similar with

\[ \lim_{k \to \infty} \frac{\mathcal{N}_k}{\mathcal{N}_{k+1}} = \lim_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = R \quad \text{and} \quad \lim_{k \to \infty} (\mathcal{N}_k R^k) = \text{const.}, \]

where \( R = 3.043827 \ldots \) and Tokunaga self-similar, asymptotically in \( k: \)

\[ \lim_{i \to \infty} T_{i,i+k} =: T_k \quad \text{and} \quad \lim_{k \to \infty} \frac{T_k}{c^{k-1}} = a \]

for some positive \( a \) and \( c. \)
Results.

**Theorem.** The asymptotic Horton ratios $\mathcal{N}_k$ exist and finite and satisfy the convergence $\lim_{k \to \infty} (\mathcal{N}_k)^{-\frac{1}{k}} = R$ with $2 \leq R \leq 4$.

Consider now a time series $X$ with $N$ local maxima separated by $N - 1$ internal local minima such that the latter form a discrete white noise; we call $X$ an *extended discrete white noise*.

**Theorem.** The combinatorial level set tree of an extended discrete white noise $X$ with $N$ local maxima has the same distribution on $\mathcal{T}_N$ as the combinatorial tree generated by Kingman’s $N$-coalescent.

The equivalence leads to the Horton self-similarity for discrete white noise.

**Corollary.** The combinatorial level set tree of a discrete white noise is root-Horton self similar with the same Horton exponent $R$ as for Kingman’s coalescent.
Horton self-similarity.

Filled circles: The asymptotic ratio $N_k$ of the number $N_k$ of branches of order $k$ to $N$ in Kingman’s coalescent, as $N \to \infty$. Black squares: The empirical ratio $N_k/N_1$ in a level-set tree for a single trajectory of a white noise of length $N = 2^{18}$. 

![Graph](image-url)
Asymptotic Tokunaga self-similarity.

Filled circles: The asymptotic Tokunaga indices $T_{i9}$ in Kingman’s coalescent, as $N \to \infty$. Black squares: The empirical Tokunaga indices averaged over 100 level-set trees for white noises of length $N = 2^{17}$.
Smoluchowski-Horton ODEs.

Let $K(i,j) \equiv 1/N$ in Kingman’s $N$-coalescent process, and let $\eta(N)(t) := |\Pi_t^{(N)}|/N$ be the total number of clusters relative to the system size $N$.

Then $\eta(N)(0) = N/N = 1$ and $\eta(N)(t)$ decreases by $1/N$ with each coalescence of clusters at the rate of

$$\frac{1}{N} \left( N \frac{\eta(N)(t)}{2} \right) = \frac{\eta^2(N)(t)}{2} \cdot N + o(N), \quad \text{as } N \to \infty$$

The limit relative number of clusters $\eta(t) = \lim_{N \to \infty} \eta(N)(t)$ satisfies the following ODE:

$$\frac{d}{dt} \eta(t) = -\frac{\eta^2(t)}{2}.$$
Smoluchowski-Horton ODEs.

For any \( j \in \mathbb{N} \) we define \( \eta_{j,N}(t) \) to be the number of clusters of Horton-Strahler order \( j \) at time \( t \) relative to the system size \( N \).

Initially,

\[
\eta_{j,N}(0) = \delta_1(j).
\]

At any time \( t \), \( \eta_{j,N}(t) \) increases by \( 1/N \) with each coalescence of clusters of Horton-Strahler order \( j - 1 \) with rate

\[
\frac{1}{N} \left( \frac{N \eta_{(j-1),N}(t)}{2} \right) = \frac{\eta_{(j-1),N}(t)^2}{2} \cdot N + o(N).
\]

Thus \( \frac{\eta_{(j-1),N}(t)^2}{2} + o(1) \) is the instantaneous rate of increase of \( \eta_{j,N}(t) \).
Smoluchowski-Horton ODEs.

For any $j \in \mathbb{N}$ we define $\eta_{j,N}(t)$ to be the number of clusters of Horton-Strahler order $j$ at time $t$ relative to the system size $N$.

Similarly, $\eta_{j,N}(t)$ decreases by $1/N$ when a cluster of order $j$ coalesces with a cluster of order strictly higher than $j$ with rate

$$
\eta_{j,N}(t) \left( \eta_{(N)}(t) - \sum_{k=1}^{j} \eta_{k,N}(t) \right) \cdot N,
$$

and it decreases by $2/N$ when a cluster of order $j$ coalesces with another cluster of order $j$ with rate

$$
\frac{1}{N} \left( \frac{N \eta_{j,N}(t)}{2} \right) = \frac{\eta_{j,N}^2(t)}{2} \cdot N + o(N).
$$

Thus the instantaneous rate of change of $\eta_{j,N}(t)$ is

$$
\eta_{j,N}(t) \left( \eta_{(N)}(t) - \sum_{k=1}^{j} \eta_{k,N}(t) \right) + \eta_{j,N}^2(t) + o(1).
$$
Smoluchowski-Horton ODEs.

Informally write the limit rates-in and the rates-out via the following Smoluchowski-Horton system of ODEs:

\[
\frac{d}{dt} \eta_j(t) = \frac{\eta_{j-1}(t)}{2} - \eta_j(t) \left( \eta(t) - \sum_{k=1}^{j-1} \eta_k(t) \right)
\]

with \( \eta_j(0) = \delta_1(j) \).

Formally, we prove hydrodynamic limit.

We show \( \eta_k(t) = \lim_{N \to \infty} \eta_{k,N}(t) \) exists, and let \( \eta_0 \equiv 0 \).

Since \( \eta_j(t) \) has the instantaneous rate of increase \( \frac{\eta_{j-1}(t)}{2} \), the relative total number of clusters of Horton-Strahler order \( j \) is given by

\[
\mathcal{N}_j = \delta_1(j) + \int_0^\infty \frac{\eta_{j-1}^2(t)}{2} dt.
\]
Smoluchowski-Horton ODEs.

**Lemma.** The Horton ratios $N_k/N$ converge in probability to a finite constant $N_k = \delta_1(k) + \int_0^\infty \frac{\eta_{k-1}^2(t)}{2} dt$ as $N \to \infty$.

$$N_1 = 1, \quad N_2 = \frac{1}{3}$$

and

$$N_3 = \frac{e^4}{128} - \frac{e^2}{8} + \frac{233}{384} = 0.109686868100941\ldots$$

Hence, we have $N_1/N_2 = 3$ and $N_2/N_3 = 3.038953879388\ldots$

Our numerical results yield, moreover,

$$\lim_{k \to \infty} (N_k)^{-\frac{1}{k}} = \lim_{k \to \infty} \frac{N_k}{N_{k+1}} = 3.0438279\ldots$$